

-1-

**Polymerase Enhancing Factor (PEF) Extracts, PEF  
Protein Complexes, Isolated PEF Protein , and  
Methods for Purifying and Identifying**

**Related Application Information**

5 This application claims priority to and is a continuation-in-part of U.S. Patent application serial no. 08/957,709, filed October 24, 1997, which is a continuation-in-part of U.S. Patent application serial no. 08/822,774, filed March 21, 1997. The entire contents of those applications are specifically incorporated herein by reference and may be relied on to make and use embodiments of the claimed invention.

**Background of the Invention**

**Field of the Invention**

The invention relates to the fields of nucleic acid polymerases and nucleic acid polymerization reactions.

15 **Introduction**

The efficiency of a nucleic acid polymerization reaction has implications for numerous assays and techniques. For example, the ability to enhance polymerase activity in a PCR process increases the sensitivity of the PCR-based assay. We have identified, produced, purified, and analyzed novel extracts, proteins, and complexes that improve the polymerization activity of nucleic acid polymerases. Included within the aspects of the present invention are methods for identifying compositions with a polymerase enhancing activity, methods for purifying and using these compositions, and specific extracts, proteins, and complexes that function to enhance polymerase activity.

25 **Description of Related Art**

30 Manipulating nucleic acids with polymerization reactions is a fundamental component of biotechnology-related research. These reactions permit researchers to replicate DNA or RNA *in vitro*, which in turn allows cloning or amplification of specific nucleic acids or groups of nucleic acids. Numerous other examples exist detailing the critical nature of a nucleic acid polymerization reaction or a nucleic acid polymerization enzyme in a particular technique, including sequencing nucleic acids, mutagenesis of nucleic acid sequences, and producing nucleic acid probes

-2-

for hybridization. Of particular current interest are amplification reactions, such as PCR, that have greatly increased the rate at which researchers can perform nucleic acid related experimentation. Extremely rare nucleic acids can now be amplified and manipulated using these techniques, which necessarily involve nucleic acid polymerases.

Using techniques with an amplification step has driven concern for the efficiency, fidelity, and sensitivity of the polymerase used. This has resulted in efforts to both analyze and optimize polymerization conditions for a variety of applications. (Lundberg et al., *Gene* 108: 1-6 (1991); Eckert and Kunkel, *PCR Methods Applic.* 1: 17-24 (1991); Ling et al., *PCR Methods Applic.* 1: 63-69 (1991); Brail et al., *Mutat. Res.* 303: 75-82 (1994); Garrity and Wold, *P.N.A.S.* 89: 1021-1025 (1992); Taylor and Logan, *Curr. Opin. Biotechnol.* 6: 24-29 (1995)) In particular, quantitative amplification-based reactions rely upon the ability to efficiently amplify each nucleic acid species present in a sample. (See Ausubel, et al., Chapter 15, In: *Current Protocols in Molecular Biology*, John Wiley & Sons (1992) and supplements through 1995.) Thus, both a concern for the accuracy of and a need for new methods to enhance the performance of amplification-based nucleic acid techniques exists in the art.

One way in which these concerns and needs have been addressed is through the use of additives to the amplification reaction. Different additives act at different points in the amplification process. For example, formamide has been used to increase the specificity of PCR with GC rich target sequences, which are particularly susceptible to intramolecular hybridization that may prevent hybridization with a primer. (Sarkar, G. et al. *Nucl. Acids Res.* 18: 7465 (1990)). It has also been reported that tetramethylammonium chloride increases yield and specificity of PCR reactions. (Chevet, E., et. al., *Nucleic Acids Res.* 23:3343-3334 (1995).) Hung et al. report the reduction in multiple satellite bands from amplifying complex DNA when dimethyl sulfoxide (DMSO) is added. (Hung, T., et al. *Nucl. Acids Res.* 18: 4953(1990).) The multiple satellite bands often present problems in purifying the desired amplification product from the other DNA present.

Certain proteins have been used to stabilize hybridized nucleic acids during replication. For example, *E. coli* single-stranded DNA binding protein has been used to increase the yield and specificity of primer extension reactions and PCR

-3-

reactions. (U. S. Patent Nos. 5,449,603 and 5,534,407.) The gene 32 protein (single stranded DNA binding protein) of phage T4 apparently improves the ability to amplify larger DNA fragments (Schwartz, et al., Nucl. Acids Res. 18: 1079 (1990)) and enhances DNA polymerase fidelity (Huang, DNA Cell. Biol. 15: 589-594 (1996)). In addition, bacterial thioredoxin combined with T7 DNA polymerase (Sequenase™; Amersham-USB) has been used to increase processivity, but the combination is not active at high temperatures, such as those used in PCR.

Another way amplification-based assays and techniques have been improved is through the development of modified polymerases or the use of combinations of polymerases. (U.S. Patent 5,566,772) For example, the TaKaRa long PCR kit employs two polymerases (Takara Shuzo Co., Ltd; Japan), and a number of polymerase combinations were also tested by Barnes (Proc. Nat. Acad. Sci. USA, 91:2216-2220 (1994). Truncated Taq and T. flavus DNA polymerase enzymes that apparently exhibit increased thermostability and fidelity in PCR have also been suggested. (U.S. Patent 5,436,149.) Combinations of polymerases with and without 5' - 3' exonuclease or 3' - 5' proofreading activity have also been used. (U.S. Patent No. 5,489,523)

Further, amplification-based assays and techniques have been improved through empirical testing of conditions, reagents, and reagent concentrations to optimize polymerization reactions with a particular enzyme. Temperature and length of amplification cycles, primer length, and pH, for example, are all conditions that can be optimized. (Barnes, Proc. Nat. Acad. Sci. USA, 91:2216-2220 (1994).)

However, accessory proteins can be even more useful in improving polymerase activity and/or the processivity of polymerases. "Processivity" in this context refers to the number of enzymatic reactions occurring each time an enzyme binds to its substrate. In the context of nucleic acid replication reactions, "processivity" means the number of bases that can be replicated when the polymerase binds to a priming site. An increase in processivity directly relates to longer replication products.

Intracellular replication has been shown to involve accessory proteins, as characterized in E. coli, human, and phage T4 systems. The accessory proteins interact with polymerases to improve activity and provide the high processivity necessary to replicate genomic DNA efficiently while avoiding unacceptable

-4-

mutation rates. Since the accessory proteins can be used in combination with the other improvements noted above, the development and application of accessory proteins holds particular promise for enhancing the results of nucleic acid replication-based reactions.

5           Accessory proteins have been identified in eukaryotes, *E. coli*, and bacteriophage-T4 and are thought to form "sliding clamp" structures. (Kelman and O'Donnell, Nucl. Acids. Res. 23(18): 3613-3620 (1995).) These structures are thought to tether the polymerase to DNA, thereby increasing processivity. The sliding clamp structures, however, have largely been studied in *in vitro* model  
10       systems. Only in the case of T4 polymerase has knowledge of the activity of such accessory proteins been used to improve polymerization-based techniques employed by researchers in the art. For example, accessory proteins of the T4 holoenzyme have been reported to improve processivity when added to polymerization systems using T4 polymerase. (Young et al., Biochem. 31(37):  
15       8675-8690 (1992); Oncor Fidelity™ Sequencing System, Oncor; Gaithersburg, MD.) However, since the T4 accessory proteins are derived from bacteriophage, they are not likely to enhance polymerases from bacteria, archae, or eukaryotes. Thus, the use of T4 accessory proteins is believed to have been limited to techniques where T4 polymerase is used.

20           The presence of dUTP (deoxyuracil triphosphate) in a polymerization reaction and the effect of deoxyuridine-containing DNA on DNA synthesis have also been examined. In particular, deoxyuridine in a DNA strand has been shown to inhibit polymerization by archael DNA polymerases. (Lasken, et al., (1996) *J. Biol. Chem.* 271; 17692-17696.) While Lasken et al. reported that archeal DNA  
25       polymerases, such as Vent, are inhibited by DNA containing deoxyuridine, they do not discuss the effect of removing uracil-containing nucleosides or nucleoside triphosphates from the reaction to prevent incorporation. Furthermore, they do not discuss any enzyme that acts on or turns over dUTP in a reaction. Neither do they mention any dUTPase activity or the possible effect of dUTPase activity on  
30       polymerization reactions. In addition, Lasken et al. do not appreciate the fact that dUTP is generated during the course of a normal PCR reaction by the deamination of dCTP. As a result of the deamination, dUTP will be present and be incorporated into an amplified nucleic acid, inhibiting the polymerase activity. Thus, the art has



-5-

not appreciated the potential of dUTPase activities and proteins in enhancing replication reactions.

Accordingly, since present knowledge and use of accessory proteins has led to limited applications in replication-based techniques, there continues to exist a need in the art for new and more widely useful compositions for enhancing polymerase enzyme activity. The present invention meets this need.

### Summary of the Invention

The present invention comprises extracts, protein complexes, and related proteins that possess nucleic acid polymerase enhancing activity useful in a variety of replication reactions known in the art. Thus, the extracts, protein complexes, and related proteins of the invention function to enhance a wide spectrum of *in vitro* nucleic acid replication reactions by providing, *inter alia*, replication products of superior length, fidelity or both, and at higher yields. As used in this specification and appended claims "polymerase enhancing activity" means the ability to increase the rate, fidelity, and/or yield of a nucleic acid polymerization reaction mediated by a nucleic acid polymerase, or to expand or alter the range of conditions under which such reaction does or may proceed.

In one aspect of the invention, extracts of *Pyrococcus furiosus* (Pfu) cells are provided that enhance the activity of Pfu DNA polymerase. The extracts enhance nucleic acid replication product yields over a fairly broad range of concentrations and contain at least one polymerase enhancing factor. As used in this specification and in the appended claims, the term "PEF" includes purified naturally occurring polymerase enhancing factors and wholly or partially synthetic copies or active analogs thereof. In accordance with the invention, such extracts can be further purified by heparin affinity chromatography followed by sepharose gel purification. Additionally, PEFs can be identified and purified using the antibodies of this invention, discussed below. While Pfu cell samples were used and are specifically exemplified below, one skilled in the art will appreciate that other cell samples can be used to identify and purify PEFs. For example, other species of the archae *Pyrococcus* or *Thermococcus* can be used as well as thermophilic bacteria cells and other bacteria cells. In addition, eukaryotic cells and tissues can be used as a source for PEF, as demonstrated by the cloning and

-6-

expression of human dUTPase, which also enhances polymerase activity. Thus, the invention also comprises compositions and methods wherein a dUTPase or any activity that turns-over dUTP is capable of acting to enhance a nucleic acid polymerization reaction.

5 In another aspect of the invention, PEF complexes are provided. The PEF complexes of the invention possess polymerase enhancing activity and generally comprise multiple protein subunits with a combined molecular weight of approximately 250kD or above as determined by SDS-PAGE analysis and gel filtration of unheated PEF samples. An example of one PEF complex (P300) was  
10 purified from Pfu cell sample extracts. The predominant components of the complex are a 50kD protein (P50) and a 45kD protein (P45). Heat treating the Pfu P45 with 2% SDS and 1% TCA produces a 17-18kD protein, which represents the fully denatured form. However, the Pfu PEF complex contains other minor components with approximate apparent molecular weights of 150, 100, 85, 60, 55,  
15 42, and 37kD. At least two components (150 and 100) have been shown to be dimeric or polymeric forms of P50. Thus, the PEF complexes of the invention comprise protein components and function to enhance the activity of polymerases.

In another aspect of the invention, Pfu proteins possessing polymerase enhancing activity are provided. These proteins have molecular weights between  
20 approximately 42 and 60kD by SDS PAGE analysis under partially denaturing conditions. The 42-60kD proteins may be used alone or in combination to enhance polymerase activity. Methods for purifying these proteins as well as the PEF extracts and PEF complexes from which they have been isolated are also provided.

The invention also involves two particular proteins, Pfu P50 and P45, which  
25 are predominant components of the PEF complex (P300). Detailed structural and functional information on the Pfu P45 and P50 proteins is disclosed. The P50 protein is similar in structure to a bacterial flavoprotein. The P45 protein is similar in structure to dCTP deaminase, functions as a dUTPase, and possesses polymerase enhancing activity. The structural information herein can be used to generate  
30 specific hybridization probes that detect the presence of nucleic acids encoding a protein that is part of a PEF complex, or related proteins from samples from other species, or possesses PEF activity. Furthermore, the structural information can be used to generate proteins from expression systems known in the art, synthetic

-7-

proteins, partially synthetic proteins, or proteins made from a combination of natural proteins, expressed proteins, and synthetic proteins. Methods for detecting the presence or absence of polymerase enhancing activity and/or dUTPase activity are also included in this invention and can be used to identify the various active PEF proteins or analogs. In addition, polyclonal or monoclonal antibodies that bind to PEF components can be produced, for example from purified P45 or P50, purified PEF complexes (P300), or another PEF of the invention. These antibodies can then be employed in assays and kits, well known in the art, in order to identify the presence or absence of a PEF.

The understanding of the catalytic activity of PEF, and the P45 protein in particular, provides aspects of this invention directed to polymerase enhancing proteins, as well as methods, kits, and compositions containing a dUTPase activity or dUTPase protein as a PEF. Thus, a dUTPase activity or dUTPase protein or composition can be used to enhance nucleic acid replication, polymerization, or PCR reactions according to this invention. In fact, any activity that functions to turn-over dUTP can be used as a polymerase enhancing activity of this invention. Wide-ranging sources for the dUTPase activity, protein, or composition exist, as it is demonstrated to be present from both archael and human sources, the ends of the phylogenetic possibilities. Thus, any cell or species can be used as a source for polymerase enhancing activity or PEF.

Kits for replicating nucleic acids and methods for using the PEF complexes, specific proteins of the complexes, and extracts containing PEF are also provided. In addition, the complexes, proteins, and extracts can be used in compositions comprising a polymerase. Ideally, the polymerase will be one that is enhanced by the complex, protein, or PEF. The PEF extracts, complexes and proteins of the present invention are particularly useful in mixtures with nucleic acid polymerases, such as native polymerases, those produced by recombinant DNA techniques, and kits containing such polymerases.

Also provided in the invention are methods for identifying proteins or complexes that influence nucleic acid polymerases. The source of the protein can be any bacterial, archael, or eukaryotic species. Certain embodiments involve methods for identifying proteins affecting polymerases used in amplification reactions, for example, alpha-type DNA polymerases such as DNA polymerases

-8-

from *Pyrococcus* and *Thermophilis* species. Other embodiments involve the analysis of dUTPase activity as well as computer implemented screening methods to identify a PEF.

#### Description of the Drawings

5           **Figure 1.** PCR enhancing activity in a heparin sepharose column fraction. PCR enhancing activity was measured using the 6.2 kb system described in example 1. Column fraction SCS #36 H.S. #78 (prep. 2) was diluted in 1 x cloned Pfu PCR buffer and 1µl aliquots of the following were added to 100µl PCRs. **Figure**  
10           **1:** lane 1, buffer; lanes 2,4 column fraction diluted 1:10; lanes 3,5 column fraction diluted 1:100. In **Figure 1**, lanes 4 and 5, λAA742 template DNA has been omitted from the reactions.

**Figure 2.** PCR enhancing activity in a heparin sepharose column fraction. **Figure 2:** lane 1, buffer; lanes 2,5 column fraction undiluted; lanes 3,6 column fraction diluted 1:10; lanes 4,7 column fraction diluted 1:100. In **Figure 2**, lanes 5-  
15           7, cloned Pfu DNA polymerase has been omitted from the PCRs.

**Figure 3.** PCR enhancing activity of SDS-PAGE gel-purified samples from heparin sepharose fraction SCS #36 H.S. #78 (prep. 2). **Figure 3** shows the PCR enhancing activity of proteins eluted from SDS-PAGE gel slices 1-7 (native Pfu DNA pol.; lanes 1-7) and 9-20 (H.S. #78; lanes 9-20 from **Figure 4**). The proteins were eluted as described in Example 2. One (1) µl of each gel slice, diluted 1:100 in cloned Pfu PCR buffer, was added to cloned Pfu PCRs as described in Example 1 (6.2kb primer-template system). In the left lanes of the gel is shown PCR product synthesis in the presence of 1µl of buffer (-) or H.S. #78, diluted 1:1000 (.001µl), 1:10,000 (.0001µl), or 1:100,000 (.00001µl). DNA markers were electrophoresed in  
20           lane "m".

**Figure 4.** **Figure 4** shows a duplicate of the master SDS-PAGE gel from which gel slices were excised. The following proteins were electrophoresed on both gels: lane A, 8 µl of native Pfu DNA polymerase (lot #24); lane B, pre-stained molecular weight markers (Novex); lane C, heparin sepharose fraction SCS #36 H.S. #78 4µl (≈ 160 ng PEF). The samples were not pre-heated before loading, and the duplicate gel shown here was silver-stained. Gel slices 1-7 were recovered from lane A on the master gel, while slices 8-25 were recovered from lane C of the master gel.  
30

**Figure 5.** DNA polymerase activity in SDS-PAGE gel purified samples. The level of DNA polymerase activity (cpms incorporated) in gel slice eluates (1 $\mu$ l) was measured as described in Example 14. The polymerase activity exhibited by gel-purified proteins numbered 1-7 in the native Pfu DNA polymerase preparation (Figure 4; lane A) is shown by the solid line (left Y axis). The polymerase activity of gel-purified proteins numbered 8-25 in fraction H.S. #78 (Figure 4, lane C) is shown with the broken line (right-handed Y axis). The apparent molecular weights of the proteins tested are shown on the x axis (at the top) and are inferred from the position the gel slices were recovered, relative to pre-stained molecular weight markers. Gel slices #11 and 18 exhibited the highest PCR enhancing activity.

**Figure 6.** SDS-PAGE analysis of gel-purified *P. furiosus* PEF. The following samples were electrophoresed on 4-20% acrylamide gels in the presence of 1% SDS: lanes 2,3- SCS #37 H.S. fraction #75 (prep. 4; 10 $\mu$ l of a Microcon 30 concentrated fraction;  $\approx$ 100 ng PEF) lanes 4,5 10 $\mu$ l of SDS-PAGE gel slice eluate recovered from SCS #37 H.S. fraction #75 (migrated 0-4mm above 250kD protein marker). The samples in lanes 3 and 5 were boiled for 3 minutes before loading onto the SDS-PAGE gel. The completed gel was stained with silver stain. The migration of pre-stained molecular weight markers is shown in lane 1.

**Figure 7.** SDS-PAGE analysis of S200-purified *P. furiosus* PEF. 10 $\mu$ l (2.25  $\mu$ g) of S200-purified PEF (prep. 1) was incubated in the presence of 2% SDS for 3 minutes at room temperature (RT), 75°C, 85°C, or 100°C (boil), as indicated on the lanes, and then subject to electrophoresis as described in Example 2. Proteins were detected by silver-staining. Protein molecular weight markers were run in lane "m".

**Figure 8.** PCR enhancing activity of SDS-PAGE gel-purified samples from an S200-purified *P. furiosus* PEF preparation (prep 1). Purified PEF ( $\approx$ 4.5  $\mu$ g) was pre-incubated in 2% SDS for 3 minutes at room temperature (RT) or at 85°C (85°C) before electrophoresis. Proteins were eluted from SDS-PAGE gel slices 1-13 (gel RT) and 14-26 (gel 85°C) as described in Example 2. One (1)  $\mu$ l of each gel slice, diluted 1:100 in cloned Pfu PCR buffer, was added to cloned Pfu PCRs as described in Example 1 (6.2kb primer-template system). The approximate molecular weights of the SDS-PAGE gel-purified proteins is indicated at the top of the gel. DNA size markers were run in lanes "m".

-10-

**Figure 9.** S200-purified PEF (1.4 µg total protein) was heated at 85 or 100°C prior to electrophoresis. Four slices were excised from the gel between the 60kD (slice 1) and 42kD (slice 4) markers. The proteins were eluted in 50µl of buffer and 10µl aliquots were boiled in the presence of SDS-BME loading dye and run out on 4-20% PAGE gels. Lane 1: Novex pre-stained markers; lane 2: 0.28 ng PEF; lanes 3-6: gel purified proteins isolated from S200-purified PEF heated at 85°C- slice 1 (lane 3), slice 2 (lane 4), slice 3 (lane 5), and slice 4 (lane 6); lanes 7-10; gel purified proteins isolated from S200-purified PEF heated at 100°C: slice 1 (lane 7), slice 2 (lane 8), slice 3 (lane 9), and slice 4 (lane 10).

**Figure 10.** SDS-PAGE analysis of *P. furiosus* PEF preparations. In the left panel, the following samples were subject to SDS-PAGE silver stain analysis, conducted as described: cloned Pfu DNA polymerase lot#24A (56 ng/µl) [lanes 1-3, 1µl, 2µl, 4µl]; S200 purified *P. furiosus* PEF fraction #46 (prep. 1 from SCS #38) [lane 4- 2µl]; S200 purified *P. furiosus* PEF pool fractions #47-48 (550 ng/µl; prep. 3 from SCS #38)[lanes 5-8, .1µl, .2µl, .4µl, 1µl]. In the right panel, the following samples were run: cloned Pfu DNA polymerase lot#24A (56 ng/µl) [lanes 1-3, 1µl, 2µl, 4µl]; microcon 30-concentrated SCS #36 heparin sepharose fraction #78 (prep. 2) [lanes 4-6, 1µl, 2µl, 0.4µl]; SCS #37 heparin sepharose fraction #75 (prep. 4) [lanes 7-8, 1µl, 5µl].

**Figure 11.** PDVF blots of *P. furiosus* PEF. PEF-containing heparin sepharose fractions (from SCS #37 Pfu purification) were concentrated and aliquots electrophoresed in 8 or 9 lanes on 4-20% SDS-PAGE gels as described. The samples were boiled for 3 min. prior to loading to recover the 50kD monomeric PEF (**top**) or were loaded in the absence of heat treatment to recover the >250kD aggregate (**bottom**). The proteins were transferred to PDVF filters (BioRad) and stained with Amido black.

**Figure 12.** PVDF blot of S200-purified *P. furiosus* PEF. Approximately 20µg of total protein was electrophoresed in each of 7 lanes on a 12% PAGE gel. The samples were heated at 85°C for 5 minutes prior to loading. The proteins were transferred to PVDF filters and stained as in the Figure 11 legend.

**Figure 13.** Absorbance spectrum of S-200 purified *P. furiosus* PEF. The spectra of the following were obtained on a Shimadzu UV160U spectrophotometer:

-11-

0.7 mg/ml *P. furiosus* PEF, 9.6 $\mu$ M and 4.8 $\mu$ M riboflavin, and PEF final storage buffer (blank).

**Figure 14.** Recombinant P45 (rP45) amplification enhancement of 5.2 kb fragment. In the 5.2 kb "On/Off" assay with PEF and rP45 samples, 7, 0.7 or 0.2 ng of the native PEF protein or 5 $\mu$ l, 1  $\mu$ l, or 1  $\mu$ l of a 1/10 dilution of the rP45 protein (20 ng/ $\mu$ l) was added to the on/off assay.

**Figure 15.** SDS-PAGE analysis of fully denatured native PEF and rP45. The following protein samples were electrophoresed on 4-20% SDS-PAGE gels as described in the text: 100 ng native PEF, boiled in 2%SDS without (lane 1) or with 1%TCA (lane 2); 200ng recombinant P45 clone 1, boiled in 2%SDS (lane 3) or loaded in the absence of heat-treatment (lane 5); 200ng recombinant P45 clone 23, boiled in 2%SDS (lane 4) or loaded in the absence of heat-treatment (lane 6). The migration of molecular weight markers is shown in lane 7 (far right). The gel was stained with SYPRO orange dye (Molecular Probes), diluted 1:5000 in 5% acetic acid.

**Figure 16.** Reverse Phase HPLC Analysis of dUTP Conversion by PEF and rP45. Panel A, dUTP heated at 72°C in 1x cloned Pfu polymerase buffer (negative control). Panel B, dUTP heated in 1x cloned Pfu polymerase buffer with 700 ng native PEF. Panel C, dUTP heated in 1x cloned Pfu polymerase buffer with 5  $\mu$ l of rP45 (clone 1), Panel D, dUMP standard heated in 1x cloned Pfu buffer, Panel E, the products shown in panel B and panel D were mixed and injected simultaneously to demonstrate, through the production of a single peak, that dUMP migrates identically to the PEF byproduct.

**Figure 17.** Reverse Phase HPLC Analysis of dCTP Conversion by PEF and rP45. Panel A, dCTP heated in 1x cloned Pfu polymerase buffer without additive (negative control), Panel B, dCTP heated in 1x cloned Pfu polymerase buffer with 700 ng of PEF. Panel C, dCTP heated in 1x cloned Pfu polymerase buffer with 5  $\mu$ l of rP45. Panel D, dUTP and dCTP heated in 1x cloned Pfu polymerase buffer with 700 ng of PEF. Chromatograms in panels A, B and D, were performed in the same series, while chromatogram C was performed in a different experiment, using a different C-18 column. In the second experiment a dCTP standard was shown to elute at 9.6 minutes (data not shown).

-12-

**Figure 18.** Reverse phase HPLC analysis of the products generated by heat-treatment of dCTP. dCTP (10mM in cloned Pfu DNA polymerase PCR buffer) was heated at 95°C in thin-walled tubes in a RoboCycler 40 temperature gradient block. Incubations were carried out for the following length of time: 0 (panel A), 1 hour (panel B), or 4 hours (panel C). 10µl of the reaction mixtures were loaded onto a Waters Delta-pak C-18 column (300 angstrom/15µm), equilibrated in 50mM triethylammonium acetate, pH(7.0) (A). Products were eluted with acetonitrile (B) at 2 ml/min., using the following gradient: 0% B for 5 minutes, 0-10% B over 20 minutes. Absorbance of the eluate was monitored with a photodiode array detector, and peak areas were integrated at 260nm. To the right of panels A and C are shown the same chromatograms re-integrated at a higher sensitivity. In panel D, 10µl of 10mM dUTP (in cloned Pfu DNA polymerase PCR buffer) was chromatographed.

**Figure 19.** Comparison of the absorbance spectrum of dCTP, dUTP, and the product generated from heating dCTP. The absorbance spectra (200-300nm) of the following were superimposed: (1) peak 9.283 from chromatogram A in Figure 18 (dCTP unheated); (2) peak 12.383 from chromatogram C in Figure 18 (heated dCTP product); (3) peak 12.100 from chromatogram D in Figure 18 (dUTP unheated). The three spectra were normalized such that heights of the maximally absorbing peaks ( $\lambda_{max}$ ) are equivalent.

**Figure 20.** Reverse phase HPLC analysis of the products generated by PCR cycling dCTP. dCTP (10mM in cloned Pfu DNA polymerase PCR buffer) was cycled alone (Panel A) or in the presence of 0.5 ng/µl PEF (Panel B) or 1.25U/µl Pfu DNA polymerase (Panel C). Cycling was carried out in a RoboCycler 40 using the following conditions: 95°C 1 min./30 cycles of: 95°C 1 min., 60°C 1 min., 72°C 10:24 min./ 72°C 10 min. 10µl of the reaction mixtures were chromatographed as described in the Figure 18 legend. In panel D, the absorbance spectra of dCTP, dUTP, and the product generated from PCR cycling dCTP are compared. The absorbance spectra (200-300nm) of the following were superimposed: (1) major peak at 8.3-8.8 min. from chromatogram of dCTP + Pfu unheated; (2) peak at 11.867 min. from chromatogram of dUTP in buffer unheated; (3) peak 12.150 from chromatogram C (Figure 20) showing dCTP + Pfu cycled.



-13-

**Figure 21. dUTP Inhibition and Reversible Inhibition with PEF and rP45.**

The following components were added to cloned Pfu DNA Polymerase PCR reactions (5.2kb "on/off" assay of example 1): Lane 1, 2  $\mu$ M dUTP + 7 ng native PEF. Lane 2, 2  $\mu$ M dUTP. Lane 3, 2  $\mu$ M PEF generated dUMP. Lane 4, 20  $\mu$ M PEF generated dUMP. Lane 5, 2  $\mu$ M dUTP + 180 ng rP45 clone 1. Lane 6, 2  $\mu$ M dUTP + 20 ng rP45 clone 1. Lane 7, 2  $\mu$ M dUTP + 2 ng rP45 clone 1. Lane 8, 2  $\mu$ M dUTP + 180 ng rP45 clone 23. Lane 9, 2  $\mu$ M dUTP + 20 ng rP45 clone 23. Lane 10, 0.2  $\mu$ M dUTP. Lane 11, 0.2  $\mu$ M dUTP + 7ng native PEF. Lane 12, 0.2  $\mu$ M dUTP + 20 ng rP45 clone 1. Lane 13, 0.2  $\mu$ M dUTP + 20 ng rP45 clone 23. In the right Lane M, 1 Kb DNA markers were run.

**Figure 22. Inhibition of PCR with unrelated dU Containing Oligonucleotides.**

A 0.9kb target was amplified from human genomic DNA. Lanes 2 and 3 represent the PCR amplification without the addition of a third oligo. The sample in lane 3 was amplified in the presence of 7ng of PEF. Lanes 4-8 and 10-11 represent the titration of the control (dT rather than dU) oligonucleotide (40mer). Lanes 13 through 32 represent samples amplified in the presence of the dU oligonucleotide (40mer). Lanes 13-19 were amplified with Pfu DNA polymerase alone. Lanes 14 through 26 were amplified with Pfu DNA polymerase in the presence of 7 ng of PEF. Lanes 27 through 32 were amplified with Taq DNA polymerase.

**Figure 23. Human dUTPase mimics PEF/P45 activity. PEF activity was measured using the 5.2kb primer-template assay described in example 1. At each annealing step of the PCR reaction, 0.5  $\mu$ l of the following were added: dUTPase storage buffer (negative control), recombinant P45 (at 2ng/ $\mu$ l)(positive control), human dUTPase preparation undiluted (duplicate reactions), or diluted 1:10 (duplicate reactions).****Figure 24. Presence of PEF(P300) in *P. furiosus* DNA polymerase preparations.**

SDS-PAGE analysis is shown for six preparations of native Pfu DNA polymerase with varying levels of PEF present. The volumes of each lot loaded is indicated ( $\mu$ l). The protein samples were not boiled prior to electrophoresis and the gels were silver-stained..

**Figure 25. Western blot analysis of crude extracts using *P. furiosus* PEF-specific IgG. Extracts were prepared from 5 different fermentations of *P. furiosus* (lanes 1-5), 3 partially purified fractions from *T. aquaticus* (lanes 6-8), and 1 extract**

-14-

from *E. coli* (lane 9). Purified PEF (550 ng) and pre-stained molecular weight markers were run in lanes 10 and 11, respectively. With the exception of the markers, all samples were boiled in SDS/BME dye prior to loading.

**Figure 26.** Western blot analysis of native PEF samples using anti-recombinant P45 antibody. Three samples (700 ng each) of native *Pfu* PEF were prepared as follows: unheated/2% SDS loading dye (lane 1); heated/2% SDS loading dye (lane 2); and heated/1%TCA/2% SDS loading dye (lane 3). Heating was carried out for 5 minutes at 95°C. The samples were electrophoresed on a 4-20% gradient gel and transferred to nitrocellulose. The blot was probed as described in the text with rabbit anti-rP45 sera. Novex prestained molecular markers (SeeBlue™) were run adjacent to the PEF samples in lane 4.

**Figure 27.** Western blot using the antibody to CBP-rP45 and cell lysates from different species. The lysates shown in lanes 2-7 were prepared by sonication. The samples were heated at 95°C prior to loading except for the sample loaded in lane 10 which was left at room temperature. Lane 1, SeeBlue™ Markers (Novex), Lane 2, Uncharacterized *Bacillus* species (Magenta), lane 3, *Thermus aquaticus*, lane 4, uncharacterized species, Lane 5, *Pyrococcus furiosus*, Lane 6, *Methanobacterium thermoautotrophicum*, lane 7, Human lymphoma HL60 cells, Lane 8 purified *Pyrococcus furiosus* PEF, Lane 9 purified fusion CBP-rP45. Lane 10 purified fusion CBP-rP45.

**Figures 28, 29, and 30.** PCR enhancing activity of *P. furiosus* PEF in cloned *Pfu* DNA polymerase PCRs. PCR amplifications were performed as described in example 14 with the following additional notes. **Figure 28.** A 1.9 kb *lacI-lacZα* target was amplified from plasmid DNA. 100μl PCRs were conducted with 50 pg of pPRIAZ, 100 ng of primers (5'CAT AGC GAA TTC GCA AAA CCT TTC GCG GTA TGG 3' (SEQ. ID NO: 20); 5'ACT ACG GAA TTC CAC GGA AAA TGC CGC TCA TCC 3' (SEQ ID NO: 21)), and 5U cloned *Pfu* DNA polymerase in the absence (duplicate samples #18) or the presence (duplicate lanes #19) of 0.5μl of a PEF-containing heparin sepharose fraction (H.S. #75; prep. 4; ≈ 10 ng/μl PEF). PCR cycling was conducted on a GeneAmp PCR System 9600 (Perkin Elmer Cetus) using the following conditions: 30s at 95°C (1 cycle)/ 5s at 95°C; 1 min. at 55°C; 2.5 min. at 72°C (30 cycles). **Figure 29.** A 10kb target from lambda DNA was amplified. 100μl PCRs were conducted with 250 ng of lambda DNA (Sigma),

-15-

250 ng of primers (F51-20 5'GGC-GTT-TCC-GTT-CTT-CTT-CG 3' (SEQ ID NO: 22) R10163-20 5'CCA-TCT-CAC-GCG-CCA-GTT-TC 3' (SEQ ID NO: 23)), and 5U cloned Pfu DNA polymerase in the absence (lane 1) or the presence of 1 $\mu$ l of a S200-purified PEF (prep. 3; 550 ng/ $\mu$ l PEF) diluted 1:500 (lane 2), 1:50 (lane 3), or 1:5 (lane 4). PCR cycling was conducted on a Robocycler 40 (Stratagene) using the following conditions: 95°C for 1 min. (1 cycle)/ 95°C for 1 min.; 62°C for 1 min.; 72°C for 10 min. (30 cycles). **Figure 30.** Lanes 1-5, a 5.2 kb portion of the human  $\alpha$ 1 antitrypsin gene was amplified from genomic DNA. 25  $\mu$ l PCRs were conducted with 62.5 ng of human genomic DNA (Promega), 50 ng of primers (F91-23 5'GAG GAG AGC AGG AAA GGT GGA AC (SEQ ID NO: 24); R5271-21 5'GCT GGG AGA AGA CTT CAC TGG) (SEQ ID NO: 25), and 0.6U cloned Pfu DNA polymerase in the absence (lane 1) or the presence of 1 $\mu$ l of S200 purified PEF (SCS #52; 0.7  $\mu$ g/ $\mu$ l) diluted 1:1000 (lane 2), 1:10,000 (lane 3), 1:100,000 (lane 4) or 1:1,000,000 (lane 5). PCR cycling was conducted on a RoboCycler 96 (Stratagene) using the following conditions: 96°C for 45s (1 cycle)/ 96°C for 45s; 60°C for 45s; 72°C for 14 min. (35 cycles)/72°C for 10 min. (1 cycle).

**Figure 31.** Effect of Stratagene's Perfect Match (*E. coli* ssb) on the specificity of PCRs conducted with cloned Pfu DNA polymerase and *P. furiosus* PEF. PCRs were conducted using the 6.2kb test system (example 1) in 100 $\mu$ l reaction volumes. 1 $\mu$ l of the following were added to PCRs: lanes 1,5-dilution buffer; lanes 2,6- undiluted PEF-containing heparin sepharose fraction (microcon 30-concentrated SCS #36 H.S. #78, prep. 2;  $\approx$ 40 ng/ $\mu$ l PEF); lanes 3,7- PEF fraction diluted 1:10; lanes 4,8- PEF fraction diluted 1:100. 1 $\mu$ l of Perfect Match was added to PCRs run in lanes 5-8. No DNA samples were loaded in the lanes between lanes 3 and 4, 4 and 5, and 7 and 8.

**Figure 32.** SDS-PAGE analysis of heparin sepharose (H.S.) column fractions. The heparin sepharose fractions indicated (# at top) from SCS native Pfu DNA polymerase prep. #37 (SCS #37) were analyzed on 4-20% SDS-PAGE gels (4 $\mu$ l/lane). Prestained molecular weight markers were run in lanes denoted "m" and 4 $\mu$ l of SCS #36 H.S. #78 (PEF prep. 2) was run in lanes marked "A". The peak of DNA polymerase (95kD) activity eluted between fractions 50 and 61.

**Figures 33, 34, 35, and 36.** Enhancing activity of *P. furiosus* PEF in PCRs conducted with Pwo (Figure 33), JDF-3 (Figures 34 and 35), ES4 and Vent (Figure

-16-

35), and Deep Vent (Figure 36) DNA polymerases. In Figure 33, 25 $\mu$ l PCRs were conducted using the 6.2kb test system (example 1) with 25U/ml Pwo DNA polymerase and 1 x cloned Pfu PCR buffer. 1 $\mu$ l of S200-purified *P. furiosus* PEF (prep 1; 225 ng/ $\mu$ l) was added undiluted (lane 1) or diluted 1:10 (lane 2), 1:1000 (lane 3), 1:10,000 (lane 4). 1 $\mu$ l of dilution buffer was added as a negative control (lane 5). In Figure 34, a 10 kb lambda DNA target was amplified from lambda DNA (lanes 1-4; 5-7; 11-13) or mouse genomic DNA, containing 40 copies (lanes 8-10) or 1 copy (lanes 14-16) of a lambda DNA transgene. 100 $\mu$ l PCRs were conducted with 250 ng of lambda DNA (Sigma) or mouse genomic DNA, and 250 ng of primers (F51-20 5'GGC-GTT-TCC-GTT-CTT-CTT-CG (SEQ ID NO: 22); R10163-20 5'CCA-TCT-CAC-GCG-CCA-GTT-TC) (SEQ ID NO: 23). PCRs were conducted in Taq PCR buffer using 5U Taq DNA polymerase (lanes 1-4) or 1U JDF3 DNA polymerase (lanes 5-16). 1 $\mu$ l of the following was added to PCRs: S200-purified PEF (prep. 3; 550 ng/ $\mu$ l PEF) diluted 1:500 (lanes 2, 12, 15), 1:100 (lanes 6,9), 1:50 (lane 3), 1:10 (lanes 7,10,13,16), or 1:5 (lane 4). 1 $\mu$ l of dilution buffer was added as a negative control (lanes 1,5,8,11,14). PCR cycling was conducted on a Robocycler40 (Stratagene) using the following conditions: 95°C for 1 min. (1 cycle)/ 95°C for 1 min.; 62°C for 1 min.; 72°C for 10 min. for Taq or 5 min. for JDF3 (30 cycles). In Figure 35, the 1.9kb ligase gene was amplified from *P. furiosus* genomic DNA. 100 $\mu$ l PCRs were conducted with 250 ng of DNA and 250 ng of primers (5'GAG CTT GCT CAA CTT TATC (SEQ ID NO: 26); 5'GAT AGA GAT AGT TTC TGG AGA CG) (SEQ ID NO: 27). PCRs were conducted with 10U ES4 DNA polymerase in Pfu PCR buffer (lanes 1,2), 1.5U JDF3 DNA polymerase in Taq PCR buffer (lanes 3,4), 4U Pfu DNA polymerase in cloned Pfu PCR buffer (lanes 5,6), 1U Vent DNA polymerase in Vent PCR buffer (lanes 7,8), or 1U Taq DNA polymerase in Taq PCR buffer (lanes 9,10). 1 $\mu$ l of the following was added to PCRs: dilution buffer (lanes 1,3,5,7,9) or S200-purified PEF (prep. 3; 550 ng/ $\mu$ l PEF) diluted 1:100 (lanes 2,4,6,8,10). PCR cycling was conducted on a DNA Thermal Cycler 480 (Perkin Elmer Cetus) using the following conditions: 95°C for 1 min.; 46°C for 1 min.; 72°C for 2 min. (30 cycles). In Figure 36, a 2kb DNA target was amplified from transgenic mouse genomic DNA. 100 $\mu$ l PCRs were conducted with 250 ng of DNA and 250 ng of primers (F51-20 5'GGC GTT TCC GTT CTT CTT CG (SEQ ID NO: 22); R2092-23 5'CGG GAT ATC GAC ATT TCT GCA CC)

-17-

(SEQ ID NO: 28). PCRs were conducted with 0.75U Deep Vent DNA polymerase in Deep Vent PCR buffer (lanes 1-4). 1µl of the following was added: dilution buffer (lane 1) or S200-purified PEF (prep. 3; 550 ng/µl PEF), diluted 1:500 (lane 2), 1:100 (lane 3), 1:50 (lane 4). PCR cycling was conducted on a Robocycler40 (Stratagene) using the following conditions: 95°C for 1 min.(1 cycle)/95°C for 1 min.; 62°C for 1 min.; 72°C for 2 min. (30 cycles).

**Figure 37.** Enhancement of RT-PCR with *P. furiosus* PEF. A portion of the EGF receptor sequence was amplified from HepG<sub>2</sub> total RNA using the 2-step RT-PCR protocol described with the following additions. 1µl of a cDNA synthesis reaction was added to 50µl PCR reactions containing 5U Pfu DNA polymerase, 1µM antisense primer (5' GAG-TTA-AAT-GCC-TAC-ACT-GTA-TCT) (SEQ ID NO: 29), 1µM sense primer [5' CAG-GAC-TCA-GAA-GCT-GCT-ATC-GAA (SEQ ID NO: 30) (1kb product) or 5'CTG-CAC-GTG-CCC-TGT-AGG-ATT-TGT (SEQ ID NO: 31) (3kb product)]. 1µl of buffer (0) or 1µl of a PEF-containing heparin sepharose fraction (SCS #37 H.S. #75; prep. 4; ≈10 ng/µl PEF) was added undiluted (UD) or diluted 1:10 to 1:10<sup>6</sup> (as indicated). PCRs were conducted in cloned Pfu PCR buffer, containing 3mM MgSO<sub>4</sub>.

**Figure 38.** Enhancement of Pfu DNA polymerase-based Seamless Cloning with *P. furiosus* PEF. 20 ng of plasmid was amplified as described in the Seamless Cloning kit protocol in the absence (lanes 2,4,6,8,10,12) or presence (duplicate lanes 1,3,5,7,9,11) of S200-purified *P. furiosus* PEF (prep. 3; 550 ng/µl), diluted 1:100. A 7.2 kb target was amplified with 6 different primer pairs (each set of 3 lanes). An extension time of 7.5 minutes was used for PEF-containing reactions, while an extension time of 15 minutes was used for reactions lacking PEF.

**Figure 39.** Enhancement of Pfu DNA polymerase-based QuikChange mutagenesis with *P. furiosus* PEF. QuikChange mutagenesis was performed using the kit control primers and plasmid template, with either native or cloned Pfu DNA polymerase. To the reactions was added 1µl of dilution buffer or 1µl of a PEF-containing heparin sepharose fraction (SCS #36 H.S. #78; prep. 2 ≈ 40 ng/µl), diluted as indicated. The relative amount of linear amplification product was assessed by the intensity of ethidium bromide-stained product bands on agarose gels. Supercompetent *E. coli* cells were transformed with the digested amplification

-18-

products. The number of amp resistant colonies and the mutation frequencies were scored.

**Figures 40 and 41.** PCR enhancing activity of S200-purified *P. furiosus* PEF. PCR enhancing activity was measured in duplicate assays using the 6.2 kb test system described in Example 1. In **Figure 40**, purified PEF (225 ng/ $\mu$ l; prep. 1) was diluted in 1 x cloned Pfu PCR buffer and 1 $\mu$ l aliquots of the following were added to 24 $\mu$ l PCRs- lane 1, buffer; lane 2, PEF undiluted; lanes 3-6, PEF diluted 1:10, 1:100, 1:1000, 1: 10,000, respectively. In **Figure 41**, the following amounts of purified PEF were added: lane 1- 1 $\mu$ l of buffer, lane 2- 1 $\mu$ l PEF neat, lanes 3-8, 1 $\mu$ l PEF diluted 1:10, 1:100, 1:1000, 1:10,000, 1:100,000, and 1:1,000,000.

**Figures 42 and 43.** PCR enhancing activity of heparin sepharose-purified *P. furiosus* PEF. PCR enhancing activity was measured using the 6.2 kb test system described. The PEF fraction ( $\approx$ 40 ng/ $\mu$ l; prep. 2 in text) was diluted in 1 x cloned Pfu PCR buffer. In **Figure 42**, 1 $\mu$ l aliquots of the following were added to 50 $\mu$ l PCRs- lane 1, buffer; lanes 2-5, PEF diluted 1:10, 1:100, 1:1000, 1: 10,000, respectively. In **Figure 43**, 1 $\mu$ l aliquots of the following were added to 24 $\mu$ l PCRs- lane 1, buffer; lane 2, PEF undiluted; lanes 3-6, PEF diluted 1:10, 1:100, 1:1000, 1: 10,000, respectively. DNA markers were electrophoresed in lane "m".

**Figure 44.** PCR enhancing activity of heparin sepharose column fractions. The PCR enhancing activity contained in 1 $\mu$ l of each column fraction (fractions 1-93 from Figure 32; numbered at top) was measured using the assay 6.2kb primer-template PCR described in Example 1. Fractions 50-61 contained the peak of Pfu DNA polymerase activity and were excluded from analysis.

#### **Detailed Description of Specific Embodiments of the Invention**

The following description should not be construed to limit the scope of this invention to any specifically described embodiment. Various aspects and embodiments of this invention will be apparent from the disclosure as a whole in context with the knowledge of one skilled in the art. In addition, the description herein, in combination with information known or available to persons of ordinary skill in the art, enables the practice of the subject matter encompassed by the following claims.

For the purposes of this invention, a nucleic acid replication reaction can mean any of the numerous nucleic acid amplification, primer extension, reverse

-19-

transcription, or other nucleic acid polymerization reactions known in the art. Additionally, a replication reaction of this invention includes any reaction in which the ability of an enzyme to interact with a first nucleic acid and generate a second, substantially complementary nucleic acid sequence, is involved. The amplification reactions of this invention are not limited to PCR processes or any particular PCR-based assay, although they are particularly useful herein, and specifically include RT-PCR processes. The proteins, preparations, compositions, mixtures, kits and methods of this invention can be used with any appropriately designed nucleic acid replication reaction.

As used herein, the term "PEF" refers to a naturally occurring protein derived from a bacterial, eukaryotic, or archael source (or a wholly or partially synthetic copy or analog thereof) having polymerase enhancing activity, mixtures of one or more such proteins, protein complexes containing one or more such proteins, or extracts containing one or more of such proteins, mixtures or complexes. Recombinant PEF proteins, as a wholly synthetic copy of a naturally occurring protein, and complexes with at least one recombinant PEF protein are also "PEFs" according to this invention. The Pfu P45 and P50 proteins of this invention are illustrative of PEF proteins, which exhibit an apparent molecular weight of approximately 45kD and 50kD and are predominant components of a PEF complex derivable from Pfu. Data relating to both the P45 and P50 proteins is presented herein and details specific structural information. On SDS-PAGE, the non-heat-treated PEF complex containing P45 and P50 as well as minor additional components migrates with an apparent molecular weight > 250kD. One species of the PEF complexes of this invention is the P300 complex from *Pyrococcus furiosus*.

The present invention is intended, however, to encompass other PEF proteins, mixtures, complexes, compositions, and extracts derived from organisms other than Pfu identified by techniques analogous to those provided by the following examples, or by use of the structural information on the PEF proteins described herein or derivable from the proteins described herein. More specifically, the invention is intended to encompass PEFs identified on the basis of sequence homology to all or part of the PEFs described herein, including nucleic acid homology to all or part of the DNA sequence encoding the proteins described herein or the DNA sequences described herein. Computer-implemented homology

-20-

searches using the sequence information herein, stored on an appropriate computer-readable medium, with procedures known in the art, can generate these homologous PEFs. Also, reactivity with antibodies to the proteins, complexes, or extracts disclosed herein can be used with procedures known in the art to generate homologous PEFs.

One skilled in the art is familiar with methods of generating analogs of proteins. Various techniques from publications in the art can be used to mutate, modify, truncate, or otherwise change a protein's amino acid sequence and retain functional activity. In the case of a dUTPase activity as PEF, the known crystal structure of dUTPases, such as *E. coli* dUTPase, provides specific information on regions of a dUTPase that may be mutated in any of a number of ways while still retaining PEF activity (see Cedergren-Zeppezauer, E. S., *et al.*, Nature 355: 740(1992), which also notes similarities to mammalian dUTPases in Hokari, S., *et al.*, Arch. Biochem. Biophys. 253: 350 (1987)). Similarly, wholly or partially synthetic or recombinantly expressed proteins can also be generated from the information herein by those skilled in the art. (For example, Ausubel *et al.* (1989) Current Protocols in Molecular Biology, and supplements through February 1997.)

Furthermore, a PEF can also be a protein exhibiting a dUTPase activity, as demonstrated herein. Specifically, human, *Pyrococcus furiosus*, and *Thermus thermophilis* dUTPase activities can be used to enhance polymerization reactions as a PEF. Other eukaryotic, bacterial, and archaeal sources can similarly be used to derive a PEF as a dUTPase protein. A PEF can also be any protein that functions to turn-over dUTP or lower the concentration of dUTP. Similarly, the invention comprises a method of enhancing replication or amplification reactions by reducing the dUTP concentration or preventing the incorporation of dUTP into replicated or amplified products, as well as compositions that are capable of preventing that incorporation.

The polymerase enhancing activity of the PEFs of this invention can be determined in a number of different ways. The description below details a few examples of assays and techniques one skilled in the art can use to determine if PEF activity is present. These assays and techniques can be used alone or in combination.



-21-

Example 1 specifically details screening assays and the "on/off" assay. This type of PCR assay allows one to identify the presence of a polymerase enhancing activity in a sample. More generally, any assay that shows an increase in PCR product yield, over a negative control level, when a sample suspected to contain a polymerase enhancing activity is added can be used to identify a polymerase enhancing activity. Also, any assay that shows an increase in processivity, over a control level, reflected by the increased length of PCR products being generated when a sample suspected to contain a polymerase enhancing activity is used. A combination of PCR product yield and increased processivity can also be used to determine whether or not a polymerase enhancing activity is present.

A polymerase enhancing activity can also be identified by assays that indicate a reduction in the PCR inhibitory action of incorporated dUTP. For example, PCR reactions can be conducted in the presence of dUTP and samples suspected of containing polymerase enhancing activity. Those reactions that allow polymerization in the presence of dUTP indicate a polymerase enhancing activity in the form of a dUTPase activity. Thus, a dUTPase activity can be a polymerase enhancing activity.

Also, a composition that functions to turn-over dUTP, especially under thermophilic reaction or PCR reaction conditions, can be a polymerase enhancing activity as a dUTPase. An enzyme or activity that acts on dUTP so that it is not incorporated into a newly polymerized strand functions to turn-over dUTP. The turn-over of dUTP can also be detected by an assay for the conversion of dUTP into dUMP, as detected by analyzing the reaction products by HPLC, for example. Biochemical assays that detect the conversion of dUTP into dUMP, or other nucleoside phosphate or metabolic derivatives or products, can be devised or are known in the art and can be used to identify polymerase enhancing activity as a dUTPase activity.

A polymerase enhancing activity can also be a dUTPase enzyme that possesses the consensus uridine-binding sequence motif (SEQ ID NO.: 72). A number of those enzymes are identified below. However, numerous others exist or can be identified through computer-implemented or other sequence analysis procedures known in the art. Thus, the presence of the consensus uridine-binding motif or the related sequences shown herein can also be used to define an enzyme

-22-

or protein that is a PEF, such as a protein that comprises SEQ ID NO.: 72, or any one of SEQ ID NOs.: 72-81, or combinations of these sequences.

Also, proteins identified through sequence identity comparisons known in the art can be used to confirm the presence of a PEF. For example, proteins from one species possessing a sequence identity of approximately 18% or greater have been shown in the art to be related to or analogous to the known protein of another species. In the examples below, a sequence similarity of approximately 39% suffices to positively identify a dUTPase activity that can act as a PEF.

The antibodies to PEF described herein can also identify a protein with polymerase enhancing activity. For example, Western blot analysis of compositions from various archeal, bacterial, thermophilic bacterial, or eukaryotic samples can identify a protein that possess polymerase enhancing activity. Furthermore, as the PEF proteins and complexes of this invention are demonstrated as immunogenic, various other antibodies to PEF may be produced by techniques known in the art with the information herein. These other antibodies can also be used to identify a PEF.

### Example 1

#### Screening for PEF Activity

Protein-containing extracts from a number of different sources can be tested for PEF activity. The extracts can be prepared in a number of ways known in the art.

One method was demonstrated with Pfu DSM 3638 cells. The cells were grown, a cell paste collected by centrifugation and then frozen at -80°C. The paste was taken up with lysis buffer [50 mM Tris-HCl (pH 8.2), 1 mM EDTA, 10 mM B-mercaptoethanol, 0.5mM PMSF, and 2µg/ml aprotinin], and thereafter the cells were lysed in a French press and then sonicated. Following sonication, the lysate was centrifuged and the supernatant, containing potential PEFs, was collected for assays.

Extracts from any cell producing a PEF, for example, cells transfected with a recombinant vector directing the expression of a PEF, can also be assayed. Methods of making extracts of these cells are known in the art and are exemplified below.

-23-

### 1. Screening Assays for PCR-Enhancing Activity

One method of detecting thermostable PEFs is by screening partially-purified fractions from thermophilic archeal or bacterial extracts for PCR enhancing activity. PCR enhancing activity can be detected in samples consisting of column-purified fractions as well as homogeneous protein samples and proteins recovered by elution from SDS-PAGE gel slices (see below). Samples are added to PCR amplification reactions containing DNA polymerase, buffer, dNTPs, primers, and DNA template. PCR enhancing activity is identified by an increase in PCR product yield for amplifications conducted in the presence of a particular sample (DNA polymerase + PEF) as compared to amplifications conducted in the absence of added sample (DNA polymerase only).

When screening samples suspected of containing endogenous DNA polymerase activity, for example protein extracts, negative controls can be performed in which the exogenous DNA polymerase has been omitted from the PCR amplifications. In addition, when screening samples contaminated with DNA, negative controls can be carried out in which exogenous DNA template is omitted from the PCR amplifications.

The sensitivity of the PCR enhancing assay is dependent on the complexity of the DNA targets employed. PCR reaction parameters (target complexity, DNA template concentration, polymerase concentration, PCR cycle number or extension time) can be adjusted so that the yield of PCR product is barely detectable under normal conditions. In addition, samples for testing can be diluted appropriately so that the concentration of PEFs falls within the detectable range of the PCR enhancing activity assay.

### 2. Screening Assays for PCR-Enhancing Activity Using Pfu-derived Samples

A number of amplification assays can be designed to detect the presence or absence of PEF activity, and/or compare PEF activity between samples. Generally, these tests employ a sample containing a rare sequence to be amplified. The sequence is so rare, or the conditions so designed, that amplification under normal situations results in barely detectable or no detectable amplified product. By adding a sample with putative PEF activity, any effects on the amount of amplified product formed can be detected.

-24-

One particular screening assay is called the "On/Off" assay, which detects the presence or absence of PEF. The "On/Off" assay results in appreciable amplified product only when PEF activity is present, or a detectable difference in amplified product when PEF activity is present compared to when PEF is not present. Methods for detecting the amount of amplified product are known in the art and include those using electrophoresis and hybridization.

One embodiment of an assay used to screen for PEFs, in this case from *Pyrococcus furiosus*, involved amplifying a 6.2 kb target from transgenic mouse DNA (6.2 kb primer-template assay). A master PCR cocktail was prepared consisting of: 200µM each dNTP, 2 µg/ml primer F432-21 (5'-CTA-TTG-AGT-ACG-AAC-GCC-ATC)(SEQ ID NO.: 62), 2 µg/ml primer R6656-20 (GTC-ACG-CTT-GCT-CCA-CTC-CG)(SEQ ID NO.: 63), 2.5 µg/ml λAA742 DNA (transgenic mouse DNA having 40 copies of a lambda phage DNA shuttle vector), 1x cloned Pfu DNA polymerase PCR buffer (10mM KCl, 10mM (NH<sub>4</sub>)<sub>2</sub>SO<sub>4</sub>, 20mM Tris-HCl (pH 8.8), 2mM MgSO<sub>4</sub>, 0.1%(v/v) Triton X-100, and 100µg/ml BSA), and 25U/ml cloned Pfu DNA polymerase. Samples were diluted in 1x cloned Pfu DNA polymerase PCR buffer, and 1µl of each diluted sample was added to 24µl of the PCR cocktail. PCR amplifications were conducted on the RoboCycler 96 Temperature Cycler (Stratagene), using the following conditions: 96°C for 45s (1 cycle)/ 96°C for 45s; 60°C for 45s; 72°C for 14 min. (35 cycles)/72°C for 10 min. (1 cycle). PCR products (10 µl/lane) are run out on 1% agarose gels and PCR product bands visualized by ethidium bromide staining. Samples with PCR enhancing activity exhibit higher yields of the expected 6.2 kb PCR product than can be obtained in the presence of DNA polymerase alone. When PCR enhancement is due to the presence of PEFs, rather than contaminating endogenous DNA polymerase activity, amplifications performed in the absence of exogenous DNA polymerase (e.g. Pfu DNA polymerase) yield no PCR product. Moreover, PCR amplifications performed in the absence of exogenous DNA template should yield no PCR product when the PEF sample lacks contaminating target DNA.

A second embodiment of an assay to screen for PEF employs, for example, the 5.2 kb human α1-antitrypsin gene in a PCR amplification. PCR amplification of this primer/template system was so limited that, in the absence of PEF, it was difficult to detect any PCR product. With added PEF activity, a 5.2 kb product was

-25-

easily detected. The following conditions were used for this "On/Off" assay: In 100  $\mu$ l - 1x Cloned Pfu DNA polymerase buffer, 200  $\mu$ M each dNTP, 200 ng primer F-91-23, 200 ng primer R5271-21, 125 ng Human Genomic DNA, 2.5 units cloned Pfu DNA polymerase, +/- PEF or recombinant P45 (rP45).

Primer F91-23      5' GAGGAGAGCAGGAAAGGTGGAAC 3'      (SEQ ID NO: 64)

Primer 5271-21      5' GCTGGGAGAAGACTTCACTGG 3'      (SEQ ID NO: 65)

The PCR cycling conditions were as follows: 95°C for 1 minute (1 cycle), 95°C for 1 minute - 60°C for 1 minute - 72°C for 10 minutes (30 cycles). After completion, the reactions are run out on an electrophoresis gel and the quantity of reaction products determined by any of a number of methods known in the art.

### 3. Screening Assays for Nucleic Acid Replication Enhancing Activity

Extracts can also be added to any nucleic acid replication reaction to determine PEF activity. Many of these reactions are known in the art, including primer extension reactions, DNA sequencing reactions, site-directed mutagenesis reactions, and a number of PCR-based reactions. (Ausubel, F.M., et al. (1989) Current Protocols in Molecular Biology, Greene Publishing Associates and Wiley-Interscience, New York, NY; Sambrook, J., et al. (1989) Molecular Cloning: A Laboratory Manual, 2nd Ed., Cold Spring Harbor Laboratory Press, Cold Spring Harbor, N.Y.) By comparing the results produced in a nucleic acid replication reaction with and without the added extract, one can identify the presence of PEF.

#### Example 2

##### Purification of PEF from *P. furiosus*

Once PCR enhancing activity has been detected from, for example, archeal or bacterial sources, large amounts of purified PEF can be obtained by column chromatography. The following protocol was developed for purifying PEF from *P. furiosus* (Pfu). However, one skilled in the art will appreciate that other cells or species could be used as well.

#### 1. Cell Growth and Lysis

*P. furiosus* DSM 3638 cells were grown in a 400 liter fermentor according to established protocol. (U.S. Patent No. 5,545,552, specifically incorporated herein by reference.) The cell paste was collected using a Sharples in-line centrifuge after approximately 20 hours ( $A_{600} \approx 0.5$ ), and then immediately frozen in liquid  $N_2$  and stored at -80°C until use. Then, 500 grams of frozen cell paste was transferred to a

-26-

4 liter stainless steel beaker on ice. The cells were resuspended with 2 liters of lysis buffer, consisting of 50 mM Tris-HCl (pH 8.2), 1 mM EDTA, 10 mM B-mercaptoethanol, 0.5mM PMSF, and 2 $\mu$ g/ml aprotinin. The cells were lysed in the French press using 2 passes at 8K PSI and the lysate was then sonicated for 10 minutes. Following sonication, the lysate was transferred to 400 ml bottles, spun for 1 hour at 9K rpm in a Sorvall RC-2B centrifuge using a Sorvall GS3 rotor, and the supernatant collected.

## 2. Purification by Column Chromatography

The supernatant was loaded at a flow rate of 5ml/min. onto a 10 x 5 cm Q-Sepharose Fast Flow™ (Pharmacia) column ( $\approx$ 392 mls), pre-equilibrated in buffer consisting of 50 mM Tris-HCl (pH 8.2), 1 mM EDTA, and 10 mM B-mercaptoethanol. The column was washed with 2 column volumes of buffer, and the pass-through and column washes were collected and pooled. The pooled fractions were adjusted to pH 7.5 using 1N HCl.

The Q-Sepharose pass-through was then loaded at a flow rate of 5 ml/min. onto a 5 x 11.5 cm ( $\approx$ 225 mls) SP Sepharose Big Bead™ (Pharmacia) column, equilibrated in buffer containing 50 mM Tris-HCl (pH 7.5), 1 mM EDTA, 1 mM DTT, 10% (v/v) glycerol, 0.1% (v/v) Igepal CA-630, and 0.1% (v/v) Tween 20. The column was washed with equilibration buffer until the absorbance (OD<sub>280</sub>) approached baseline. The column was eluted with a 2 liter gradient from 0 to 250 mM KCl (in equilibration buffer). Fractions of 20 ml were collected, and aliquots removed from every third tube for SDS-PAGE analysis.

Some of the fractions analyzed by SDS-PAGE showed a band > 250kD when a sample was not heated prior to electrophoresis ( $\approx$ 300kD). The fractions containing the 300 kD band were pooled and dialyzed overnight against 2 x 4 liters of Buffer A [50 mM Tris-HCl (pH 8.2), 1 mM EDTA, 1 mM DTT, 10% (v/v) glycerol, 0.1% (v/v) Igepal CA-630, and 0.1% (v/v) Tween 20]. The dialyzed pool was loaded at a flow rate of 2 ml/min. onto a 2.6 x 29 cm ( $\approx$ 154 mls) Heparin Sepharose CL-6B™ (Pharmacia) column, equilibrated in Buffer A. The column was washed with 1 liter of Buffer A, and then eluted with a 1.5 liter gradient from 0 to 300 mM KCl/Buffer A. Fractions of 10 ml were collected, and aliquots removed from every third tube for SDS-PAGE analysis. Fractions containing the 300kD band were pooled and dialyzed overnight against 2 x 4 liters of Buffer A.

-27-

The heparin sepharose-purified pool was loaded at a flow rate of 0.5 ml/min. onto a 1.6 x 95 cm ( $\approx$ 191 mls) Sephacryl S-200 High Resolution™ (Pharmacia) column equilibrated in Buffer A containing 100 mM KCl. Then, 2 ml fractions were collected and aliquots removed from every third tube for SDS-PAGE analysis. Fractions containing the 300kD band were pooled and dialyzed overnight against 1 liter of buffer containing 50 mM Tris-HCl (pH 8.2), 0.1 mM EDTA, 1 mM DTT, 50% (v/v) glycerol, 0.1% (v/v) Igepal CA-630, and 0.1% (v/v) Tween 20. The purified protein was stored at -20°C. The purification protocol described above yielded  $\approx$ 1 mg. of relatively homogeneous P300 band from 500 g. of cell paste.

### 3. Purification of PEF from SDS-PAGE gels

PEF in a heterogeneous sample can be identified by eluting purified protein from SDS-PAGE gel slices and rescreening for PCR enhancing activity. This method allows rapid assessment of the number of PEF proteins in a particular sample and identification of their apparent molecular weight.

*P. furiosus* protein samples with PCR enhancing activity were electrophoresed on 4-20% acrylamide/2.6% bis-acrylamide Tris-Glycine gels (Novex), along-side pre-stained molecular weight markers. Samples were loaded in the presence of 2% SDS, but were not boiled in order to prevent dissociation of PEF complexes. The gels were run in Tris-Glycine buffer containing 1% SDS, and after electrophoresis, were washed briefly in 20mM Tris-HCl (pH 8.0)/1mM EDTA. Then, 2-3mm gel slices were excised from the lanes of interest using the pre-stained molecular weight markers as a guide. Each gel slice was cut-up with a razor blade and the pieces transferred to 50 $\mu$ l of elution buffer (20mM Tris-HCl (pH 8.0)/1mM EDTA/0.1% Tween-20). The slurry was incubated at 72°C for 30 minutes.

Gel slices containing PEFs are identified by testing the eluates for PCR enhancing activity. Eluates containing  $>0.1$  ng/ $\mu$ l PEF are then re-analyzed on silver-stained SDS-PAGE gels to verify the apparent molecular weight of the predominant protein component. The gel slice eluates are boiled in the presence of 2% SDS before loading and the apparent molecular weights of PEF proteins determined relative to protein standards. The gel slice elution procedure described here allows recovery of 1-10% of the protein of interest.

-28-

**Example 3****Identification of the PCR Enhancing Activity in *P. furiosus* Partially-Purified Column Fractions**

5 The fractions collected after the heparin sepharose chromatography were analyzed for PEF activity using the PCR screening assay (Example 1). The addition of diluted heparin sepharose fraction dramatically increased yields of PCR products generated with cloned Pfu DNA polymerase. The PCR enhancing activity of the fractions was shown to be dependent upon the presence of exogenous DNA template (Figure 1) and Pfu DNA polymerase (Figure 2). Increased PCR product yield was, therefore, not due to the presence of contaminating DNA template or native Pfu DNA polymerase, but rather to the presence of PEFs.

10 In order to further characterize the factor or factors responsible, the following was performed. PEFs after heparin sepharose chromatography were identified by screening SDS-PAGE gel-purified samples for PCR enhancing activity, as discussed above in Example 2. When the protein samples were loaded onto SDS-PAGE gels without pre-boiling, PCR enhancing activity (Figure 3) was recovered in 2 gel slices from the gel of Figure 4. One gel slice (gel slice #1) was excised from a position between the 42 and 60kD markers, while the second gel slice (slice #2) was recovered from a site just above the 250kD marker (Figure 4).

15 The proteins eluted from the gel slices were also screened for DNA polymerase activity to demonstrate that PCR enhancing activity was not related to contaminating DNA polymerase activity (Figure 5). The results indicated that SDS-PAGE purified proteins with PCR enhancing activity lack significant DNA polymerase activity. Moreover, SDS-PAGE purified Pfu DNA polymerase lacks PCR enhancing activity when tested at protein concentrations comparable to or greater than those of gel-purified PEFs.

20 The *P. furiosus* PEF proteins were conclusively identified by re-analyzing gel purified samples with PCR enhancing activity on silver stained SDS-PAGE gels. In the absence of pre-boiling, the predominant band in gel slice #2 migrates somewhat slower than the 250kD molecular weight marker, consistent with the site where the gel slice was recovered (Figure 6). This band is called P300 or PEF complex. In addition, a minor band is evident at 50kD, called P50. Gel slice #1 contained a major band at 50kD and a minor band at 45kD.



-29-

However, when the proteins eluted from gel slice #2 are boiled in SDS prior to loading, the predominant component migrates with an apparent molecular weight of 50kD (Figure 6). A minor or poorly staining component of  $\approx 45$ kD is also visible. These results are consistent with *P. furiosus* PEF consisting of two distinct proteins, with apparent molecular weights of 50kD and 45kD, which aggregate at low temperatures in presence of SDS to produce a complex which migrates as a discrete band at 300kD.

#### Example 4

##### Characterization of PEF Complex and Protein Components of the Complex

*P. furiosus* PEF fractions following S200 chromatography comprises a mixture of proteins. A discrete band in SDS-PAGE migrates above the 250kD marker when the sample is not heated prior to electrophoresis. This protein is called P300 or PEF complex. When the conditions are changed to substantially dissociate the PEF complex, the subunit protein components of the PEF complex are visualized in SDS-PAGE.

##### 1. Temperature-dependent Dissociation of PEF Complex

One method of dissociating PEF complex into its component proteins is heat treatment. In the absence of heating, the majority of PEF migrates as a complex, running slightly slower than the 250kD molecular weight marker. Minor amounts of dissociated P50 and P45 are visible in the unheated sample. However, after heat treatment at temperatures of about 85°C or higher, the PEF complex is completely dissociated as indicated by the absence of the 300kD band by SDS-PAGE. The predominant protein component of PEF complex, as visualized by silver-staining, exhibits an apparent molecular weight of approximately 50kD in SDS-PAGE. The P50 band is shown in the gel of Figure 7, where the protein was heated to  $\geq 85^\circ\text{C}$  in the presence of 2% SDS and BME prior to loading. In addition, the P45 protein shown to be present in the gel-purified PEF complex (Figure 6) is also evident in the heat-treated, S200-purified PEF sample (Figure 7). Furthermore, minor components with apparent molecular weights of approximately 37, 42, 55, 60, 85, 100, and 150kD were also detected in SDS-PAGE analyses of PEF complex treated at temperatures of 85-100°C. These minor components may represent additional forms of P50 and P45 generated by heat treatment (e.g. dimers, trimers, fragments) or minor unrelated species.

-30-

## 2. Polymerase Enhancing Activity of PEF Complex and Component Proteins and Mixtures

Protein components of S200-purified PEF were purified by SDS-PAGE and the PCR enhancing titer was determined by adding serial dilutions of each gel slice eluate to PCR reactions with cloned Pfu DNA polymerase PCR reactions. The protein or protein mixtures which exhibited the highest levels of polymerase enhancing activity were identified by running the eluates on silver stained SDS-PAGE gels. Analysis of SDS-PAGE gel slice eluates indicates that PCR enhancing activity of S200-purified *P. furiosus* PEF can be attributed solely to the 45kD species plus the 50kD species. When an S200-purified preparation was loaded in the absence of heating (Figure 7, prep. 1), PCR enhancing activity was present in gel slices recovered just above the 250kD marker and between the 42 and 60kD markers. When heated to 85°C before loading, PCR enhancing activity migrated between the 42 and 60kD markers (Figure 8).

Titration experiments showed that the PCR enhancing titer of gel purified proteins migrating with apparent mass between 42kD and 60kD was not significantly different from that of the gel-purified PEF complex (300kD band). The levels of PCR enhancing activity migrating between the 148 and 60kD markers were insignificant, thereby indicating that the 85kD, 100kD, and 150kD bands do not contribute substantially to full PCR enhancing activity of the PEF complex (P300). Moreover, these components do not appear to further enhance the activity of PEF.

Figure 9 shows the proteins recovered from 4 gel slices between the 42 and 60kD markers from 2 heated treated PEF samples. The greatest PCR enhancements were observed for protein samples run in lanes 5, 9, and 10. These lanes contained the highest amounts of P45, in addition to low but detectable amounts of P50. Relative P50 concentration did not necessarily correlate with highest PCR enhancing titer, as the proteins run in lane 4 (where only P50 is visible) exhibited a lower titer than the protein mixture in lane 5 (same amount of P50, plus P45). Moreover, samples in lanes 9 and 10 exhibited similarly high PCR enhancing titers and levels of P45, but the sample in lane 9 contained 10 to 1000-fold more P50 than the sample in lane 10. These results are consistent with P45 being the most active component of the PEF complex. Since all samples of gel

-31-

purified P45 isolated contain varying concentrations of P50, the exact biochemical role a low concentration of P50 has on attaining full PEF activity or stability could not be determined by this method.

#### Example 5

##### Amino Acid Analysis of PEF Complex and P50 and P45 Components

The complex and the predominate 50kD component (P50) and 45kD (P45) component from Pfu were sequenced at the N-terminus. In addition, N-terminal sequence analysis was performed on the minor 100 and 150kD components generated upon heat dissociation. Two analyses were performed. In the first study, heparin sepharose-purified PEF samples ( $\approx 20\%$  homogeneous; prep. 4 in Figure 10) were electrophoresed and electroblotted onto PDVF filters. Samples were loaded onto 4-20% SDS-PAGE gels at room temperature or after heat treatment, to allow recovery of both the 50kD protein and the  $>250$ kD complex (Figure 11). Blots were sent to Wistar Sequencing Facility (Philadelphia, PA) for analysis. N-terminal sequence analysis was performed on both the 50kD (P50) protein (heated SDS-PAGE sample) and the  $>250$ kD PEF complex (unheated SDS-PAGE sample).

The N-terminal amino acid sequences of the PEF complex (P300) and the 50kD component (Pfu P50) were found to be substantially identical (Table A). This data confirmed that Pfu P50 is the predominant component of the PEF complex. Two distinct sequences were found for both P300 and P50 ( $1^\circ$  and  $2^\circ$ ), suggesting that Pfu PEF may contain 2 different 50kD species which co-migrate, or that the PVDF strip containing the 50kD species was contaminated with the 45kD species or other species visible by SDS-PAGE.

In addition to the N-terminal sequencing, the 50kD protein was also subject to *in situ* trypsin digestion and microbore reverse HPLC. A subset of tryptic peptides was analyzed by mass spec. Two peptides with single masses (#107, #112) and one peptide with two masses (#108) were chosen for sequence analysis. Two internal peptide sequences from Pfu P50 were recovered (Table A; Tryptic Peptides). Peptide #112 was 24 amino acids in length and the calculated mass of the Edman sequence (2530.8) was in very good agreement with the observed peptide mass (2531.7). Peptides #107 and #108 contained multiple sequences which could not be sorted by Edman sequencing alone. However these peptide

-32-

fractions eluted very close together on microbore HPLC and contained several residues in common. Based upon shared sequence and mass analysis, a tentative sequence was assigned (107/108; Table A).

### Table A - Analysis 1.

#### N-terminal Sequencing

Protein	Mass (Da)		Sequence/(SEQ ID NO: )	
Sample	Obs.			
PEF complex	≈50,000	1°	<u>XL</u> LHH <u>VK</u> LIY <u>AT</u> XX <u>B</u>	(1)
		2°	XXXPD WXXRX EXLXX	(2)
P50	≈50,000	1°	<u>ML</u> LHH <u>VK</u> LIY <u>AT</u> K <u>S</u> R	(3)
			<u>RL</u> VGK <u>KI</u> VLA <u>IP</u> G <u>XI</u> <u>AA</u> VEP	
		2°	XXXPD <u>WS</u> XXRX EXLGE <u>KFY</u>	(4)

#### Tryptic Peptides

Peptide	Mass (Da)		Sequence/(SEQ ID NO: )	
Comments	Obs.	Calc.		
107	1389.59	N/A		
multiple sequences	108	1659.1, 1910.63	N/A	
107/108	-	1910.3	<u>KY</u> DAV IMAAA VVD <u>F</u> R PK	(5)
AAs common to 107/108				
112	2531.73	2530.8	ADLVV GNTLE AFGSE	(6)
			ENQVV LIGR	

"X" represents any amino acid

underlined residues represent amino acids that may be deleted or substituted with any amino acid but are tentatively assigned as indicated

The 35 amino acid sequence recovered from the N-terminus of Pfu P50 (SEQ ID NO.: 3), and the two internal peptides of 17 and 24 amino acids (SEQ ID NO.:s 5 and 6), represent approximately 16% of the total amino acid sequence of Pfu P50, assuming an apparent molecular weight of 50kD and a length of approximately 454 amino acids.

In the second round of analyses, the N-terminal sequences of the 150, 100, 50, 45, and 42kD species were determined from a PVDF blot of heated S200-purified PEF (Figure 12 and Table B). This analysis was performed by the Beckman Research Institute of the City of Hope (Duarte, CA). The N-terminal sequences of the 150 and 100kD species were the same and identical to the major sequence in the 50kD band (except for ambiguity at the N-terminus) and similar to the minor sequence in the 45kD band. In addition, the sequence was very similar to

-33-

the major sequence in the PEF complex and the 50kD band determined in the Wistar sequence analysis. The data are consistent with the 150 and 100kD species being alternative forms of the 50kD species (e.g., dimers, trimers, or aggregates). The major N-terminal sequence of the 45kD band (2 blot sections analyzed; "upper" and "lower") was distinct from the P50 sequence and very similar to the minor sequence found in the 50kD bands analyzed by both Wistar and Beckman and in the PEF complex analyzed by Wistar. No N-terminal sequence was recovered for the 42kD species. In total, these data are consistent with the PEF complex consisting of 2 distinct protein components, P50 and P45.

**Table B - Analysis 2.**

Protein sample		Sequence	
(SEQ ID NO)			
(MW)			
150kD		(GAM)LHHV KLIYA <u>T</u> KLRK	(7)
100kD		(GAM)LHHV KLIYA TK( <u>KL</u> )RK	(8)
50	1°	M LHHV KLIYA TKL	(9)
	2°	<u>GL</u> ( <u>KL</u> )PD W( <u>WK</u> )(KF)RK EES	(10)
45 (upper)	1°	( <u>GAI</u> )LLPD WKIRK EILIE	(11)
	2°	XMHH(VI) KLXYA TXSRK	(12)
45 (lower)	1°	M(LY) (LV) (RP)D WKRRK EILIE	(13)
42	no sequence		

X represents any amino acid;

underlined residues represent amino acids that may be substituted with any amino acid but are assigned as indicated;

amino acids within parenthesis ( ) indicate that one of the enclosed amino acids is present at that site.

Using sequence information stored in a computer readable medium, one skilled in the art can perform computer-implemented homology searches. Here, the nonredundant GenBankCDS translations+PDB+SwissProt+SPupdate+PIR protein databases, using BLASTp, indicated that the partial amino acid sequence of Pfu P50 and P45 do not exhibit identity to any protein in those databases.

**Example 6****Nucleotide and Predicted Amino Acid Sequence of P50**

The nucleotide sequence of the P50 protein component was obtained by cloning the Pfu P50 using standard techniques.

-34-

### 1. Library Screening

A *Pyrococcus furiosus* genomic library was plated on XL1-Blue MRF' *E. coli* at a density of approximately 2000 plaques per plate. Duralose filters (nitrocellulose on a nylon backing) were used to take replicate lifts from each plate. While the first filter was on the plate, orientation marks were made by stabbing a needle through the filter and into the plate. The orientation marks were marked in pen on the back of the plate before the filter was removed. The filter lifts were treated as follows:

1.5-2.0 minutes	1.5 M NaCl, 0.5M NaOH
2 minutes	0.5 M Tris (pH 8.0), 1.5 M NaCl
30 seconds	2xSSC, 0.2M Tris (pH 7.5)

After treatment, the filters were partially dried until they were still damp, but no standing water was visible. The DNA on the filters was fixed by UV crosslinking with the Stratalinker set to the "Autolink" format.

The filters were prehybridized in 15 ml of:

5x SSC  
40 mM NaPO<sub>4</sub> pH (6.5)  
5x Denhardt's  
5% Dextran Sulfate  
50% Formamide

0.1 mg/ml Salmon sperm DNA (Boiled separately and added immediately prior to use)

Prehybridization was carried out at 42°C for approximately 2 hours.

Probe was generated from the 900bp PCR product amplified from Pfu genomic DNA with the following degenerate primers:

Oligo #50: CAT CAT GAA AAA CTA ATT TAC GC (SEQ ID NO: 14)  
                  C    C       C    G T C  
                              T       T

Oligo #61: GC CAT AAT TAC TGC ATC GTA TTT (SEQ ID NO:15)  
                  G    C    G    G  
                  T    G    A  
                              A

Oligo #50 was designed to hybridize to DNA encoding the HHVKLIYA (SEQ ID NO.: 66) peptide in SEQ ID NO.: 1, at the N-terminus of P50, while oligo #61 was designed to hybridize to the antisense DNA strand encoding the peptide KYDAVIMA (SEQ ID NO.: 67) in SEQ ID NO.: 5.

-35-

The PCR product was purified from free primers, buffer and nucleotides and 50 ng was labeled with  $^{32}\text{P}$ - $\alpha\text{dATP}$  using the Stratagene Prime-It II Random Primer Labelling kit. The probe was purified from free nucleotides before being boiled for five minutes and added to the prehybridization reaction. The total probe was roughly calculated to be 80 million cpm.

Hybridization was allowed to continue overnight at 42°C before the hybridization solution was removed and the filters were washed four times with 0.1x SSC, 0.1% SDS at 60°C (stringent conditions).

The filters were exposed to X-ray film overnight and 20 primary isolates, with strong signals on both replicate filters, were picked.

Six primary isolates were diluted, plated, and screened again using the same method described above. Of the six, three filters produced isolated lambda clones. The clones were confirmed by PCR amplification using the degenerate primers. All clones were able to produce the 900 bp amplified product with oligos #50 and #61, which was used as probe. Clones 6A and 3B produced a 1200 bp amplified fragment with oligos #54 and #58. Oligo #54 was designed to hybridize to DNA encoding the HHVKLIYA (SEQ ID NO: 66) peptide in SEQ ID NO: 1, and oligo #58 was designed to hybridize to the antisense DNA strand encoding the EENQVVL (SEQ ID NO: 68) peptide of SEQ ID NO: 6. Clone 6D only produced a 900 bp amplified product.

Oligo #54:           CAT CAT GAA AAA CTA ATA TAC GC (SEQ ID NO: 16)  
                          C    C       C       G T C  
                                  T           T

Oligo #58:           AG TAC TAC TTG ATT TTC TTC (SEQ ID NO: 17)  
                          A G    G    C    G  
                          A    A

Bluescript plasmid was excised from the lambda clones in SOLR cells and the presence of inserts confirmed again by PCR amplification of the 1200 or 900bp product.

## 2. DNA Sequencing

Sequencing was carried out on purified PCR products and plasmid mini-preps made from the excised cells. The nucleotide sequence is listed below with the predicted protein translation. The peptide sequences used to generate the probes are indicated by underlining. "N" represents any base and "X" represents any amino acid.

-36-

(SEQ ID NO: 18)

ATGCTTCACCACGTCAAGCTAATCTACGCCACAAAAAGTCGAAAGCTAGT  
 TGGAAAAAGATAGTCNNNNNNNNNCCAGGGAGTATTGCGGCTTTGGATG  
 TGAAAGCTTGTGAGGGACTAATTAGGCATGGGGCCGAAGTTCATGCAGTG  
 5 ATGAGTGAGGCAGCCACCAAGATAATTCATCCTTATGCATGGAATTTGCC  
 CACGGGAAATCCAGTCATAACTGAGATCACTGGATTTATCGAGCATGTTG  
 AGTTAGCAGGGGAACATGAGAATAAAGCAGATTTAATTTTGGTTTGTCTT  
 GCCACTGCCAACACAATTAGTAAGATTGCATGTGGAATAGATGATACTCC  
 10 AGTAACTACAGTCGTGACCACAGCATTTCACACATTCCAATTATGATAG  
 CCCCAGCAATGCATGAGACAATGTACAGGCATCCCATAGTAAGGGAGAAC  
 ATTGAAAGGTTAAAGAAGCTTGGCGTTGAGTTTATAGGACCAAGAATTGA  
 GGAGGGAAAGGCAAAAGTTGCAAGCATTGATGAAATAGTTTACAGAGTTA  
 TTA AAAAGCTCCACAAAAAACATTGGAAGGGAAGAGAGTCCTAGTAACG  
 15 GCGGGAGCAACAAGAGAGTACATAGATCCAATAAGATTCAACAAATGC  
 CAGCAGTGGAAAAATGGGAGTAGCGTTGGCTGAAGAAGCAGATTTTAGAG  
 GAGCTGTTACCCTCATAAGAACAAGGGAAGTGTAAAGGCTTTTAGAATC  
 AGAAAAATCAAATTGAAGGTTGAGACAGTGAAGAAATGCTTTCAGCGAT  
 TGAAATGAGTTGAGGAGTAAAAAGTATGACGTAGTTATTATGGCAGCTG  
 CTGTAAGCGATTTTAGGCCAAAAATTAAGCAGAGGGAAAAATTAAGC  
 20 GGAAGATCAATAACGATAGAGCTCGTTCCNNNNAATCCCAAAATCATTGA  
 TAGAATAAAGGAAATCAACCAATGTCTTTCTTGTGGATTTAAAGCAG  
 AAACCTCAAAGAAAGCTTATAGAAGAAGGTAAAGGCAGATTGAGAGG  
 GCCAAGGCTGACTTAGTCGTTGGTAACACATTGGAAGCCTTTGGAAGCGA  
 GGAAACCAAGTAGTATTAATTGGCAGAGATTTACAAAAAGAACTTCAA  
 25 AAATGAAAAAGAGAGATTAGCAGAGAGAATTTGGGATGAGATAGAGAAA  
 TTNCTGTCC

*Pyrococcus furiosus* P50 predicted amino acid sequence:

(SEQ ID NO: 19)

MLHHVKLIYATKSRKLVGKKIVXXXPGSIAALDVKACEGLIRHGAEVHAV  
 30 MSEAA TKI IHPYAWNLP TGNPVITEITGFIEHVELAGEHENKADLILVCP  
 ATANTISKIACGIDDPVTTTVTTAFPHIPIMIA PAMHETMYRHPIVREN  
 IERLKKLGVEFIGPRIEEGKAKVASIDEIVYRVIKKLHKKTLEGKRVLT  
 AGATREYIDPIRFITNASSGKM GVALAEADFRGAVTLIRTKGSVKAFRI  
 35 RKIKLKVETVEEMLSAIENELRSKKYDVVIMAAAVSDFRPK IKAEGKIKS  
 GRSITIELVPXNPKIIDRIKEIQPNVFLVGFKAETSKEK LIEEGKRQIER  
 AKADLVVGNTLEAFGSEENQVVLIGR DFTKELPKMKKRELAERIWDEIEK  
 XLS

Translated sequence corresponding to chemically-determined N-terminal sequence (SEQ ID NO.: 3):

MLHHVKLIYATKSRKLVGKKIVXXXPGSIAA (SEQ ID NO: 46)

Translated sequences corresponding chemically-determined internal peptide sequences (SEQ ID NOs.: 5 and 6):

KYDVVIMAAAVSDFRPK (SEQ ID NO: 47)

ADLVVGNTLEAFGSEENQVVLIGR (SEQ ID NO: 48)

The protein has a theoretical pI of 9.36 and a theoretical MW of 44801.29.

There are inconsistencies between the chemically-determined AA sequence of P50 and the AA sequence derived from the *P. furiosus* P50 genomic clone. One skilled in the art is familiar with many reasons for this type of inconsistency. For example, the inconsistencies below can, largely, be explained by known limitations common to the procedures used. These limitations do not operate to limit the



-37-

structural knowledge of proteins or nucleic acids. Instead, they merely indicate possible variations in the sequences of amino acids or nucleic acids by a finite number.

Some of the inconsistencies and explanations are:

5 MLLHHVKLIYA TKSRR LVGKK IVLAI PGXIA AVEP (Table A; SEQ ID NO:s 1 and 3)

MLHHV KLIYA TKL (Table B; SEQ ID NO: 9)

MLHHV KLIYA TKSRR LVGKK IVLAI PGSIA ALDV (predicted sequence from SEQ ID NO.:19)

10 The inconsistency in Table A sequence at cycle 2 (extra AA between AA1 and 2) may be due to contamination with P45, which appears to have L's at positions 2 and 3. Moreover, L at cycle 2 in SEQ ID NO: 1 was assigned tentatively. Other inconsistencies between the Table A sequence and the predicted sequence occur at AA 15 (R vs. K) and AA 32-34 (VEP vs. LDV).

15 An inconsistency between the Table B sequence and the predicted sequence was found at AA13. The identification of AA13 as L instead of S is explained by the poor recovery of S in chemical sequencing and the contamination of P50 with low amounts of P45, which has a L at that position.

20 ADLVV GNTLE AFGSE ENQVV LIGR (Table A; SEQ ID NO: 6)  
ADLVV GNTLE AFGSE ENQVV LIGR (predicted sequence from  
SEQ ID NO.: 19)

KYDAV IMAAA VVDFR PK (Table A; SEQ ID NO: 5)  
KYDVV IMAAA VSDFR PK (predicted sequence from  
SEQ ID NO.:19)

25 SEQ ID NO: 6, determined chemically from a P50 tryptic peptide, was identical to a 24 AA sequence translated from the P50 DNA sequence. For SEQ ID NO: 5, there were 2 inconsistencies found between the chemical and DNA sequences. An A was recovered at cycle 4 instead of a V, and a V was recovered at cycle 12 instead of a S. The inconsistencies may be due to the difficulties associated with  
30 interpreting sequences from a sample that is not absolutely pure.

#### Example 7

#### Identification of Proteins Related to P50: Similarity to *E.coli* DFP Flavoprotein

##### 1. DNA Sequence Similarity of Pfu P50 Protein to Archaeal and Bacterial Proteins

35 The DNA sequence of a P50 clone exhibits very strong homology to the flavoprotein DFP, a protein identified in *E. coli* as playing a role in DNA synthesis

-38-

and pantothenate metabolism (Spitzer and Weiss, J. Bacteriol. 164:994-1003 (1985) and Spitzer, Jimenez-Billini, and Weiss, J. Bacteriol. 170:872-876 (1988)). Although DFP was found to be an essential gene for DNA replication, these authors were not able to elucidate its role in DNA replication. The sequences in Table 1 (N-terminal 1° sequence, tryptic peptides 107/108 and 112) are all found in the translated P50 clone, which exhibits very high similarity to DFP. Accordingly, the P50 amino acid and DNA sequence information can be used to identify related proteins associated with PEF complexes from other sources such as bacteria, eukaryotes, and archae.

The amino acid sequences of *Methanococcus jannaschii* (Mja) and *E. coli* dfp proteins support the identification of the protein designated P50 (Pfu) as a member of the dfp family of proteins. The three protein sequences were compared using ClustalW 1.6, with the comparison data represented below.

	1	15	16	30	31	45	46	60	61	75	76	90	
1 Pfu	-----MLHH	VKLIYATKSRKLVGK	KIVXXXPGSIAALDV	-KACEGLIRHGAEVH	AVMSEAATKIIHPYA	WNLPTGNPVIETITG	78						
2 Mja	-----MISEIMHP	TKLLKGTISKLLLENK	KILVAVTSSIAAIET	PKLMRELIRHGAEVY	CIITEETKKIIIGKEA	LKFGCGNEVEEITG	83						
3 E.coli	MKARQOKYCDKIANF	WCHPTGKIIHSLAGK	KIVLGVSGGIAAYKT	PELVRRRLDRGADVR	VAHTEAAKAFITPLS	LQAVSGYPVSDSLD	90						
Page 2.1													
1 Pfu	91	105	106	120	121	135	136	150	151	165	166	180	
2 Mja	-----FIEHVELAGE	HENKADLIIVCPATA	NTISKIACGIDDPV	TTVTTTAFPHIPIMI	APAMHETMYRHPVR	ENIERLK-KLGVEFI	162						
3 E.coli	-----DIEHILY--	---NECDCLLIYPATA	NIISKINLGIADNIV	NTTALMFFGNKPIFI	VPAMHENMFN--AIK	RHIDKLKEKDIYII	162						
Page 3.1													
1 Pfu	181	195	196	210	211	225	226	240	241	255	256	270	
2 Mja	GPRIIE-----GKA	KVASIDEIVRVVKK	LHKKTLE-GKRVLV	AGATREYIDPIRFT	NASSGKMGVALAEEA	DFRGAVTLIRTKGSV	245						
3 E.coli	SPKFEE-----GKA	KVANIEDVVKAVIEK	IGNHLKKEGNRVLL	NGGTVEFIDKVRVIS	NLSSGKMGVALAEAF	CKEGPFYEVITAMGL	246						
Page 4.1													
1 Pfu	271	285	286	300	301	315	316	330	331	345	346	360	
2 Mja	KAFRIKIKLVETV	EHLMSAIENELRSKK	YDVVIHMAAVSDFRP	KIKAEKGKISGRS--	--ITIELVPXNPKII	DRIKEIOPN-VFLVG	330						
3 E.coli	EPPIYIKNHKVLTAK	EHLNKAIE--L-AKD	FDIIISAAISDFTV	ES-FEGKLSSEEE--	--LILKLKR-NPKVL	EELRRIYKD-KVIIG	326						
Page 5.1													
1 Pfu	361	375	376	390	391	405	406	420	421	435	436	450	
2 Mja	FKAETSK-EK-LIEE	GKROIERAKADLVVG	NTL-----EAFGSEEN	QVVLIGRDFTKELPK	MKKRELAERINDEIE	KXLS-----	403						
3 E.coli	FKAETNN-----VEEY	AKERLNKYNLNMIIA	NDLSK--HYFGDDYI	EVYIITKYEVEKISG	SKK-EISERIVEKVK	KLVKS----	403						
												430	

From the above comparison, it would be apparent to one of skill in the art that related proteins from other species can be identified and isolated by methods known in the art. The example above employed stringent screening conditions. Less stringent conditions, varying the concentration of salts, detergent, or the temperature during hybridization or washing, as known in the art, would lead to related clones from libraries containing sequences of any of a number of species. For example, in addition to the conditions described above, any of the following hybridization conditions can be used, in any combination, in methods to isolate DNA sequences related to the P50 or P45 sequences herein:

-39-

low stringency wash in a solution comprising approx. .45 M NaCl, approx. .045 M trisodium citrate, and approx. .1% SDS, at approx. 37° to approx. 42°C;

hybridization buffer comprising approx. .75M NaCl, approx. .15 M Tris, approx. 10 mM sodium pyrophosphate, approx. .075 M trisodium citrate, and approx. 50% formamide;

hybridization buffer comprising approx. 5 x SSC, approx. 5x Denhardt's, approx. 5% Dextran Sulfate, approx. 50% formamide, and approx. .1 mg/ml ssDNA;

hybridization wash comprising approx. .1 M phosphate, approx. .1 x SET, approx. .1% sodium pyrophosphate, and approx. .1% SDS at approx. 45°C.

## **2. Absorbance Spectrum of Purified *P. furiosus* PEF Complex**

The absorbance spectrum of purified *P. furiosus* PEF complex reveals two peaks of absorbance at 370 and 450 nm. Figure 13 depicts the absorbance spectrum of S-200 purified *P. furiosus* PEF. These data indicate and are consistent with the PEF complex comprising at least one flavoprotein. Sequencing data also verifies the identification of *P. furiosus* P50 as a homolog of the *E. coli* DFP protein. *E. coli* DFP is a flavoprotein containing a non-covalently associated FMN moiety.

Up to this point, flavoproteins have not been directly implicated as part of the replication machinery. The potential involvement of a flavoprotein in PCR enhancement suggests a role for redox reactions. The only redox reaction involved in DNA synthesis is the formation of deoxyribonucleotides from ribonucleotides, which is catalyzed by ribonucleoside diphosphate reductase. *In vitro*, the ribonucleoside diphosphate reductase enzyme can be coupled to NADPH via two known pathways involving FAD-containing oxidoreductases (Pigiet and Conley, J. Biol. Chem. 252:6367-72 (1977); Thelander and Reichard, Ann. Rev. Biochem. 48:133-158 (1979)). One pathway involves thioredoxin and thioredoxin reductase. Interestingly, *E. coli* thioredoxin has been shown to act as an accessory protein and confer processivity on T7 DNA polymerase. An alternate hypothesis for the potential role of a flavoprotein in PCR enhancement is that a flavoprotein may be required for the continuous processing or activation of other proteins or cofactors needed for nucleic acid replication.

-40-

**Example 8****Nucleotide and Amino Acid Sequence of P45**

The nucleotide sequence of the Pfu P45 protein component was obtained as described below.

**1. Synthesis of P45 Sequences**

Amino terminal peptide sequencing of purified P45 protein allowed the generation of four degenerate oligonucleotides designed to hybridize to DNA encoding the PDWKIRKE (SEQ ID NO.:69) peptide of SEQ ID NO.: 11, as follows:

#743: CCA GAC TGG AAA ATA AGG AAA GA (SEQ ID NO: 32)  
           T                  G TGG  
                               C  
                               T

#744: CCA GAC TGG AAA ATA AGA AAA GA (SEQ ID NO: 33)  
           T                  G TGG  
                               C  
                               T

#745: CCA GAC TGG AAA ATA AGG AAG GA (SEQ ID NO: 34)  
           T                  G TGG  
                               C  
                               T

#746: CCA GAC TGG AAA ATA AGA AAG GA (SEQ ID NO: 35)  
           T                  G TGG  
                               C  
                               T

A lambda phage *P. furiosus* genomic library was used as template for PCR amplification of the P45 sequence. The four degenerate oligonucleotides were used in separate reactions to prime template in one direction, in combination with one of the primers that border the genomic insertion of the lambda vector (-20 primer and reverse primer). The PCR reaction was carried out as specified below:

10 µl 10x Stratagene cloned *Pfu* buffer

5 µl degenerate P45 primer (either # 743, 744, 745 or 746) at 100 ng/µl

2.0 µl either reverse or -20 primer (100 ng/µl)

0.8 µl 100 mM dNTP (total dNTPs)

0.5 µl *Taq* DNA polymerase (Stratagene, 5u/µl)

0.5 µl *Taq* Extender (Stratagene, 5u/µl)

3.0 µl *Pfu* genomic library (~1.2 x 10<sup>10</sup> plaque forming units/ml)

78.2 µl H<sub>2</sub>O

-41-

PCR cycling was carried out in a RoboGradient temperature cycler as follows: One cycle at 95°C for 3 minutes, followed by 30 cycles of: 95°C for 1 minute; 51°C to 65°C gradient for 2 minutes; 72°C for 6 minutes.

The PCR products were separated on a 1% agarose, 1x TBE gel. All primer combinations produced multiple bands. A pattern of four bands was consistently seen with primers 743, 744, and 746 in conjunction with the -20 primer. The three degenerate primers that formed consistent four band patterns with the -20 primer were able to generate the pattern at 56°C. Only primer 743 could generate the pattern at 58°C. The band pattern produced with the degenerate primers in combination with the reverse primers was less distinct and formed only at lower annealing temperatures than the products generated with the -20 primer.

## 2. Cloning Strategies

Two strategies were used to isolate the P45 clone. One procedure was to make simplified sub-libraries of the original highly complex library and screen for an insert with the -20 and 743 primers. Positive sub-libraries could be diluted and rescreened until individual plaques containing the appropriate insert were identified. The other technique was to make use of Vectorette™ technology (Genosys Biotechnologies), which allows PCR amplification when the sequence of only one end of a DNA fragment is known. In the vectorette system, genomic DNA is digested with a selection of specific restriction endonucleases. After digestion, the ends of the genomic DNA are ligated to specific vectorette units, which have the same cohesive termini as the genomic DNA digestion. The ligated vectorette unit contains a sequence complimentary to a provided vectorette PCR primer. (Arnold and Hodgson, PCR Methods and Applications 1: 39-42 (1991).)

## 3. The Vectorette Reaction

Fifty µl reactions containing 100 ng of *P. furiosus* genomic DNA were digested with *Eco* RI, *Hind* III and *Bsp* 106I (an isoschizomer for *Cla* I) in their recommended buffers for one hour at 37°C. Without any post-reaction treatment, 1 µl of the appropriate vectorette unit (*Hind* III, *Cla* I or *Eco* RI at 3 pmole/µl) was added with 6 µl of 10 mM ATP, 1 µl of 100 mM DTT and 1 unit of T4 DNA ligase (Stratagene 4 u/µl). The reaction was cycled at the following temperatures: 20°C for 60 minute followed by 37°C for 30 minutes for 3 cycles.

-42-

The ligated DNA was amplified according to the following:

10 µl cloned 10x Stratagene *Pfu* buffer

8.3 µl degenerate P45 primer at 100 ng/µl

2.0 µl 50 pmol/µl vectorette primer

5 0.8 µl 100 mM (total) dNTP

0.5 µl *Taq* DNA polymerase (Stratagene, 5u/µl)

0.5 µl *Taq* Extender (Stratagene, 5u/µl)

1.0 µl vectorette library

76.9 µl H<sub>2</sub>O

10 PCR cycling was carried out as follows: One cycle at 95°C for 1 minute followed by 30 cycles of: 95°C for 1 minute; 56°C for 2 minutes; and 72°C for 3 minutes.

Ten µl were loaded on an 1% agarose, 1x TBE gel. Multiple bands were produced by all primers except 745. To determine if all three vectorette library products had been correctly primed off the same target DNA (P45 sequence) rather than having been produced by a non-specific PCR reaction, the products were  
15 digested with *Mnl* I. *Mnl* I cleaves at a frequent four base pair recognition sequence and produces a useful pattern of bands specific to the template digested. The pattern generated by electrophoresis of the *Mnl* I digestion fragments of the *Cla* I/743, *Hind* III/744 and *Eco* RI/744 PCR products on a 6% acrylamide gel showed  
20 some variation, but the majority of bands could be identified in all three samples, indicating that they share large segments of identical sequence.

#### 4. Screening

The PCR products from the *Cla* I/743 and *Hind* III/744 combinations were mixed and purified from free nucleotides and unused primers before being used as  
25 template for the generation of a 52 million cpm <sup>32</sup>P labeled probe. Details on probe synthesis and library screening are cited in Example 6.

More than 60 positive clones resulted from screening with the mixed vectorette probe. Several positive were well situated for collection without significant contamination from adjoining plaques. Twelve of these plaques were  
30 subjected to PCR amplification with the 743 and -20 primer as described previously except that an annealing temperature of 56°C was used instead of a temperature gradient. In the same amplification assay, 11 sub-libraries were assayed in the same manner.

-43-

Three of the twelve clones recovered from the primary radioactive label screen produced strong, single bands. Clone 1 produced a band of approximately 5 kb, clone 3 produced a band of approximately 3.5 kb, and clone 9 generated a band of approximately 2.7 kb. One of the sub-libraries also produced a clone of approximately 6.5 kb.

## 5. Sequencing

Sequencing of the P45 clones was carried out on purified PCR products and plasmid mini-preps made from excised cells. The nucleotide sequence of P45 is listed below with the predicted amino acid translation. The chemically-determined N-terminal sequence of P45 (SEQ ID NO.: 11), used to generate the degenerate PCR primers (SEQ ID NO.: 32-35), is indicated by underlining.

### P45 nucleotide sequence (SEQ ID NO.: 70)

ATGCTACTTC CAGACTGGAA AATCAGAAAA GAAATACTTA TAGAGCCATT  
 TTCTGAAGAA TCGCTCCAAC CAGCAGGTTA TGACCTCAGA GTGGGCAGAG  
 AGGCTTTTGT TAAGGGGAAA TTAATCGACG TGGAAAAGGA AGGAAAAGTC  
 GTTATTCCCTC CAAGGGAATA CGCCTTAATC CTAACCCTCG AGAGGATAAA  
 GTTGCCCGAC GATGTTATGG GGGATATGAA GATAAGGAGC AGTTTAGCAA  
 GAGAAGGGGT TATTGGTTCT TTTGCTTGGG TTGACCCAGG ATGGGATGGA  
 AACTTAACAC TAATGCTCTA CAATGCCTCA AATGAACCTG TCGAATTAAG  
 ATATGGAGAG AGATTTGTGC AGATCGCATT TATAAGGCTA GAGGGTCCGG  
 CAAGAAACCC TTACAGAGGA AACTATCAGG GGAGCACAAAG GTTAGCGTTT  
 TCAAAGAGAA AGAAACTCTA G

### P45 amino acid sequence (SEQ ID NO.: 71)

MLLPDWKIRK EILIEPFSEE SLQAGYDLR VGREAFVKGK LIDVEKEGKV  
 VIPPREYALI LTLERIKLPD DVMGDMKIRS SLAREGVIGS FAWVDPGWDG  
 NLTLMLYNAS NEPVELRYGE RFVQIAFIRL EGPARNPYRG NYQGSTRLAF  
 SKRKKL\*

The translated P45 protein has a theoretical pI of 9.12 and a calculated molecular weight of 17868.76. The translated N-terminal sequence (underlined) of P45 corresponds to the experimentally-determined sequence (SEQ ID NO.: 11) and matches the exact sequence (SEQ ID NO.: 60) used to design the degenerate PCR primers.

When the P45 DNA sequence is translated in all six frames and compared to multiple sequence databases using the computer-implemented program Blastx, the dCTP deaminase gene of *Desulfurolobus ambivalens* was found to have similarities. Another entirely different gene encoding polyprotein from Visna and Maeda/Visna virus was also identified, but at a less significant level of sequence

-44-

similarity. Each of these similar genes, as well as those discussed below, may represent sequences related to P45. Thus, sequence information stored on computer readable medium may be used with computer based methods to search for homologous, similar, or identical sequences in a database to develop further P45 and PEF-encoding DNA sequences in accordance with this invention, as shown in example 9 below.

### Example 9

#### Identification of Proteins Related to P45

##### 1. DNA Sequence Homology of *P. furiosus* P45 Protein to dCTP Deaminase and dUTPases

When the P45 DNA sequence was compared to multiple databases using the program BlastX, the probable deoxycytidine triphosphate deaminase (dCTP deaminase) gene (*dcd*) of *Desulfurolobus ambivalens* was found to exhibit the highest degree of similarity. The biochemical properties and physiological role of dCTP deaminase has been studied in *E. coli*, where dCTP deaminase is a homotetramer, which catalyzes the formation of dUTP and NH<sub>3</sub> from dCTP. DNA sequences with similarity to the *E. coli dcd* gene have been discovered in other bacteria (*Salmonella typhimurium*, *Haemophilus influenzae*) and in archaea (Ouzounis, C., Kyrpides, N., and Sander, C. Nucl. Acids Res. 23:565-570 (1995); Bult, C.J. et al. Science 273:1058-1073 (1996); Beck, C.J., Eisenhardt, A.R. and Neuhard, J., J. Biol. Chem. 250:609-616 (1975); Fleischmann, R.D. et al., Science 269:496-512(1995)). Amino acid sequence comparisons between *P. furiosus* P45 and the dCTP deaminases from *Desulfurolobus ambivalens* ("A") (also known as *Acidianus ambivalens*, *Sulfolobus ambivalens*), *E. coli* ("E"), and *Haemophilus influenza* are shown below, demonstrating sequence similarity ranges from about 39.1 to about 71.2%.



-45-

MAP Multiple Sequence Alignment Results									
Page 1.1	1	15 16	30 31	45 46	60 61	75 76	90		
1 p45	MLLPDMKIRKE---	ILIEPFSE-EWLQ	AGYDLRVGXX	XRVR	-----EAXVKGK---	LIDVEK-----	EGKVKXIPPREYAL		
65									
2 A.	MILGDRDLKYYLEKG	WIVISPLTO-DTIRE	NGVDLRVGG--EIA	R FKTYDEIYEDGKDPR	SFYIEIK-----	--GDEFIIPNEHVL	77		
3 E.	MRLCDRDIEAWLDEG	RLSINPRPPVERING	ATVDVRLGNKFRTER	GHTAAFIDLSGPKDE	VSAALDRVMSDEIVL	DEGEAFYLHPGELAL	90		
4 HAEIN	MRLCDTDIERVLDG	IISLTTPRPNNDKING	ATIDVRLGNSFRVFR	EHSAPFIDLSGPKDE	VSAQLESVMSDEIII	PEGEAFTLHPGTLAL	90		
Page 2.1	91	105 106	120 121	135 136	150 151	165 166	180		
1 p45	ILTLERIKLPDDVNG	IMKIRSSLAAREGVIG	SFAW--VDPGWDGNL	TLMLYNASNEPVELR	YGERFVOIAFIRLEG	PARNPYR----	GNYO	149	
2 A.	LVTEEYVKLPNDVMA	FVNLRSSFARLGLFV	PPTI--VDAGFEGOL	TIEVLG-SAFPVKIK	RGTRFLHLIFARTLT	PVENPYH----	GKYO	160	
3 E.	AVTLESVTLPADLVG	WLDGRSSLARLGLMV	HVTAHRIDPGWGGCI	VLEFYNSGKLPALR	PQMLIGALSFEPLSG	PAVRPYNRRERAKYR	180		
4 HAEIN	ATTLESVKLPANIIG	WLDGRSSLARLGLMV	HVTAHRIDPGWGGKI	VLEFYNSGKLPALR	PQMLIGALSFEVLG	EXKRPYSSRDKAKYK	180		
Page 3.1	181	195 196	210 211	225 226	240 241	255 256	270		
1 p45	GSTRLAFSKRKKL--	162							
2 A.	GOOGVTLPKFKFR--	173							
3 E.	NOOGAVASRIDKD--	193							
4 HAEIN	NOOGAVASRIDEDKE	195							

E. coli DCD exhibits an apparent molecular weight of 21.2kD (Wang, L. and Weiss, B. J. Bacteriol. 174:5647-5653 (1992)), while the predicted molecular weight of *M. jannaschii* DCD is approximately 22kD (204 amino acids). These molecular weights are approximately half the apparent molecular weight of *P. furiosus* P45 and suggest that the heat-dissociated form of P45 (apparent mol. weight of 45kD) may actually be a dimer.

In addition to dCTP deaminase, *P. furiosus* P45 exhibits a significant, but lower degree of sequence similarity to uridine triphosphatase (dUTPase). dUTPase, an enzyme encoded by the *dut* gene, converts dUTP to pyrophosphate and dUMP. An amino sequence alignment comparing the sequence of P45 to several dUTPases is shown below. Regions of identity are shown in black boxes, while regions of similarity are shown in grey boxes.



-47-

One of the regions of sequence similarity between P45 and dUTPase is the putative uridine-binding motif. This motif is conserved in the translated amino acid sequence of *P. furiosus* P45, which reveals the presence of a putative uridine-binding sequence conserved in  $\psi$  synthetases, dCTP deaminases, and dUTPases (Koonin, E.V. Nucl. Acids Res. 24:2411-2415 (1996)). In the following comparisons, U represents a bulky hydrophobic residue such as I, L, V, M, F, Y, or W, and the bolded residues match the G, D or U residues of the consensus sequence.

Consensus uridine-binding

motif:	.GUUD..U.G.U.U	(SEQ ID NO.: 72)
<i>P. fur.</i> P45:	FAWVDPGWDGNTLM	(SEQ ID NO.: 73)
<i>M. jann.</i> DCD:	AGWIDAGFKGKITL	(SEQ ID NO.: 74)
<i>M. jann put.</i> DCD:	SAVHDPGYEGRPEY	(SEQ ID NO.: 75)
<i>D. sulf.</i> DCD:	PTIVDAGFEGQLTI	(SEQ ID NO.: 76)
<i>E. coli</i> DCD:	AHRIDPGWSGCIVL	(SEQ ID NO.: 77)
<i>E. coli</i> DUT:	VGLIDSDYQGQLMI	(SEQ ID NO.: 78)
Yeast DUT:	AGVVDRDYTGEVKV	(SEQ ID NO.: 79)
Human DUT:	AGVIDEDYRGNVGV	(SEQ ID NO.: 80)
Herpes virus DUT:	TGLIDPGFQGELKL	(SEQ ID NO.: 81)

Each of these proteins represent activities, such as dUTPase, that may be used as a polymerase enhancing activity or PEF. One skilled in the art can identify numerous other proteins using stored sequence information, in the appropriate computer readable medium, from this disclosure and analogous searching procedures in other databases. Database information on each of the following species can specifically be used to identify PEF using one or more of the sequences, or parts thereof, identified herein: *Pyrococcus furiosus*; *Pyrococcus horikoshii*; *Pyrobaculum aerophilum*; *Sulfolobus solfataricus*; *Archaeoglobus fulgidus*; *Aquifex pyrophilus*; *Methanococcus jannaschii*; *Thermotoga maritima*; *Methanobacterium thermoautotrophicum*; and *Thermus thermophilis*.

The physiological function of dCTP deaminase has only been studied in *E. coli*, where it plays an essential role in deoxyribonucleotide metabolism. dCTP deaminase converts dCTP to dUTP, which is an obligatory step in the *de novo* synthesis of thymidylate in bacteria (Taylor, A.F., and Weiss, B. J. Bacteriol. 151:351-357 (1982)). In turn, uridine triphosphatase (dUTPase encoded by the *dut* gene), a ubiquitous enzyme found in bacteria, eukaryotes, and eukaryotic viruses, degrades dUTP to pyrophosphate and dUMP, the thymidylate synthetase substrate.

-48-

Thus, dCTP deaminase and dUTPase are functionally linked, with mutations in the *dcd* gene suppressing *dut* mutations (Wang, L. and Weiss, B. J. Bacteriol. 174: 5647-5653 (1992)).

dUTPase has shown to be an essential gene in *E. coli* and in yeast (El-Hajj, H.H., Zhang, H., and Weiss, B. J. Bacteriol. 170: 1069-1075 (1988); Gadsden, M.H., et al. EMBO J. 12:4425-4431 (1993)) because it functions *in vivo* to prevent dUTP incorporation into DNA. In *E. coli dut* mutants, the dUTP pool is greatly enhanced, resulting in an increased substitution of uracil for thymine in DNA during replication. Uracil-DNA glycosylase and exonuclease III play an essential role in repairing uracil-containing DNA in *E. coli dut* mutants (Taylor, A.F. and Weiss, B., J. Bacteriol. 151:351-357 (1982)).

The substrate specificities, enzyme activities, and physiological role of dCTP deaminase and dUTPase had not been characterized in archae prior to this disclosure.

#### Example 10

##### Expression Cloning and Characterization of Recombinant P45

##### 1. Expression Cloning of P45 by the Method of Ligation Independent Cloning (LIC)

Recombinant P45 was produced by PCR amplification of a portion of a P45 genomic clone (clone #9). The primers (oligos # 1 and 2 below) were designed to function with the Affinity Protein Expression and Purification System (Stratagene; La Jolla, CA), which uses Ligation Independent Cloning (LIC).

oligo # 1. 5' **GACGACGACAAGAT**GTGCTACTTCCAGACTGGAAA 3' (SEQ ID NO: 82)

oligo # 2. 5' **GGAACAAGACCCGT**CCCACTTTCACAGATGAAGAG 3' (SEQ ID NO: 83)

The bold letter segments represent sequences specific to the cloning vector while the adjoining sequence is specific to the clone #9 sequence. The ATG underlined in oligo #1 corresponds precisely to the 5' end of the P45 gene, while the sequence after the bold letters in oligo #2 corresponds to the end of the genomic insert.

The PCR amplification was carried out in a volume of 100 µl containing: 1x Cloned Pfu Polymerase Buffer; 0.2 mM dNTPs (each); 200 ng of Primer oligo #1; 200 ng of Primer oligo #2; 3 µl Genomic clone #9 plaque core in 500 µl SM buffer (~2000 Lambda particles); 2.5 units Cloned Pfu DNA Polymerase; and 7 ng Native

-49-

PEF (where 10x Cloned Pfu Polymerase Buffer is: 100 mM KCl; 100 mM  $(\text{NH}_4)_2\text{SO}_4$ ; 200 mM Tris-Cl (pH 8.75); 20 mM  $\text{MgSO}_4$ ; 1% Triton® X-100; and 1000  $\mu\text{g/ml}$  BSA).

The thermocycling parameters were: 95°C for 1 minute (1 cycle); 95°C for 1 minute - 56°C for 1 minute - 72°C for 5 minutes (30 cycles).

The 2.5kb amplified product was purified and then subjected to limited nucleotide excision in the presence of dATP. This protocol promotes removal of nucleotides at the 3' termini of the PCR product until an adenine residue is reached. The excision mixture (consisting of: 1x Cloned Pfu Polymerase Buffer; 0.5 mM dATP; 43.5  $\mu\text{l}$  PCR product (8.7 ng/ $\mu\text{l}$ ); 1.25 units Cloned Pfu DNA polymerase) was incubated at 72°C for 10 minutes.

20  $\mu\text{l}$  of the exonuclease treated PCR product was annealed with 40 ng of digested pCAL-n-EK vector (exonuclease treated to produce ends complimentary to the sequence exposed in the PCR product) for 45 minutes at room temperature. The amount of insert molar excess, relative to vector, was approximately 9 fold. The pCAL-n-EK vector contains an upstream, in-frame calmodulin binding peptide tag, which allows the N-terminal fusion protein to be easily purified on calmodulin agarose (CAM agarose). Various other expression vectors, which may or may not produce fusion proteins, are known in the art and can be used to express P45 protein or fragments thereof or to produce DNA constructs with a sequence encoding P45 protein or fragments thereof. An expression vector need only contain DNA sequences operating to permit or control transcription from an appropriately linked nucleic acid. The type of control, the degree of transcription permitted, and the manner in which the vector and nucleic acid are appropriately linked may vary. Generally, an expression vector also contains a replication control sequence to allow the vector to replicate in a host. However, replication control sequences are not required where replication of the host is not crucial to expression.

## 2. Purification of Recombinant P45

Five microliters of the annealed vector/insert DNA was transformed into XL2-Blue competent cells. Ten of the resultant colonies were screened by PCR for the 2.5 kb insert and 9 were found to contain the correct size insert. Plasmid DNA was prepared from three clones and then used to transform BL21(DE3) cells. Six BL21(DE3) colonies were grown for approximately 10 hours in 5 ml of 1x LB, 125

-50-

µg/ml ampicillin at 37°C. These cultures were used to inoculate six flasks containing 250 ml 1x LB and 125 µg/ml ampicillin. When the optical density (OD<sub>600</sub> values) of the cultures reached 1.1 to 1.3, IPTG was added to a final concentration of about 1 mM. The cells were grown overnight at 37°C. Cells were collected by centrifugation and the pellet estimated (visually) to be in the range 0.4 - 0.5 gm. The pellet was resuspended in 2.5 ml of calcium binding buffer: 50 mM Tris-HCl (pH 8.0); 150 mM NaCl; 10 mM β-ME; 10 mM magnesium acetate; 2 mM CaCl<sub>2</sub>.

250 µl of 10 mg/ml lysozyme was added to the cells and the reaction was allowed to incubate on ice for one hour. The slightly lysed samples were sonicated twice with the Branson Sonifier 250, the microtip at a duty cycle of 50% and a setting of 4. The lysate was cleared by superspeed centrifugation. Cleared lysate was added to 700 µl of washed Stratagene Calmodulin agarose (50% CAM agarose by volume) and allowed to bind with gentle rocking at 4°C for 1 hour. The resin was washed 3 times with 10 ml of CaCl<sub>2</sub> binding buffer and twice with 5 ml of the same solution except that the CaCl<sub>2</sub> concentration was reduced to 0.1 mM. Washing was accomplished by centrifugation, removal of the supernatant, and resuspension in fresh buffer. 5 µl of the CAM resin was collected to examine on SDS-PAGE electrophoresis. The washed matrix was resuspended in 900 µl of elution buffer [50 mM Tris-HCl (pH 8.0); 150 mM NaCl; 10 mM β-ME; 2 mM EGTA] and allowed to sit for one minute prior to pelleting of the agarose resin. The elution buffer containing P45 was removed to a separate tube and the CAM agarose was resuspended again in 900 µl of elution buffer. Next, a high salt elution buffer [50 mM Tris-HCl (pH 8.0); 1000 mM NaCl; 10 mM β-ME; 10 mM magnesium acetate; 2 mM EGTA] was used to elute remaining P45 protein from the CAM agarose:

Subsequent SDS-PAGE analyses showed that the high salt elution buffer released a majority of the recombinant *P. furiosus* P45 protein from the column matrices. Some residual protein remained bound to the calmodulin agarose.

The method described here to produce P45 protein can be modified in numerous ways by methods known in the art. (Ausubel, F.M., et al. (1989) Current Protocols in Molecular Biology, Greene Publishing Associates and Wiley-Interscience, New York, NY; Sambrook, J., et al. (1989) Molecular Cloning: A Laboratory Manual, 2nd Ed., Cold Spring Harbor Laboratory Press, Cold Spring Harbor, N.Y.) In one possible modification, a P45 analog protein can be produced.

-51-

For example, a mutation can be introduced into the P45 coding region. Any type of mutation can be used including site-specific point mutation, deletion mutation, insertion mutation, and multiples or combinations of these mutations. This mutant coding region is inserted into an appropriate vector, which is transferred into a host cell. The host cell then expresses the P45 analog. A P45 analog protein substantially retains one or more of the PEF activity or dUTP or dCTP metabolic activities described herein. Thus, the fusion protein and affinity tag expression and purification system described here is only one of many ways to produce a recombinant PEF protein such as recombinant P45.

Analogs may be constructed by, for example, making various substitutions of residues or sequences or deleting terminal or internal residues or sequences not needed for biological activity. For example, cysteine residues can be deleted or replaced with other amino acids to prevent formation of unnecessary or incorrect intramolecular disulfide bridges upon renaturation. Other approaches to mutagenesis involve modification of adjacent dibasic amino acid residues to enhance expression in yeast systems in which KEX2 protease activity is present. Other mutations can be made that favor expression in various host cells.

Typically, substitutions may be made conservatively. For example, one may substitute amino acids that have physiochemical characteristics resembling those of the residue to be replaced. Similarly, when a deletion or insertion strategy is adopted, the potential effect of the deletion or insertion on biological activity should be considered. In order to preserve the biological activity, deletions and substitutions will preferably result in homologous or conservatively substituted sequences, meaning that a given residue is replaced by a biologically similar residue. Examples of conservative substitutions include, but are not limited to, substitution of one aliphatic residue for another, such as Ile, Val, Leu, or Ala for one another, or substitutions of one polar residue for another, such as between Lys and Arg; Glu and Asp; or Gln and Asn. In view of this specification, one skilled in the art will be able to make other such conservative substitutions, for example, substitutions of entire regions having similar hydrophobicity characteristics

Mutations in nucleotide sequences constructed for expression of an analog, in most instances, should preserve the reading frame phase of the coding sequences and preferably will not create complementary regions that could hybridize to produce secondary mRNA structures such as loops or hairpins which would

-52-

adversely affect translation of the receptor mRNA. Although a mutation site may be predetermined, it is not necessary that the nature of the mutation per se be predetermined. For example, in order to select for optimum characteristics of mutants at a given site, random mutagenesis may be conducted at the target codon and the expressed mutants or analogs screened for the desired activity.

Not all mutations in the nucleotide sequence which encode the protein will be expressed in the final product. For example, nucleotide substitutions may be made to enhance expression, primarily to avoid secondary structure loops in the transcribed mRNA (see EPA 75,444A, incorporated herein by reference), or to provide codons that are more readily translated by the selected host, e.g., the well-known *E. coli* preference codons for *E. coli* expression.

Mutations can be introduced at particular loci by synthesizing oligonucleotides containing a mutant sequence, flanked by restriction sites enabling ligation to fragments of the native sequence. Following ligation, the resulting reconstructed sequence encodes an analog having the desired amino acid insertion, substitution, or deletion.

Alternatively, oligonucleotide-directed site-specific mutagenesis procedures can be employed to provide an altered gene having particular codons altered according to the substitution, deletion, or insertion required. Exemplary methods of making the alterations set forth above are disclosed by Walder et al. (Gene 42:133, 1986); Bauer et al. (Gene 37:73, 1985); Craik (BioTechniques, Jan. 12-19, 1985); Smith et al. (Genetic Engineering: Principles and Methods, Plenum Press, 1981); and U.S. Pat. Nos. 4,518,584 and 4,737,462, which disclose suitable techniques, and are incorporated by reference herein.

### **3. Comparison of Recombinant P45 and Native PEF PCR Enhancing Activity**

The 5.2kb "On/Off" assay, described in example 1, was used to demonstrate that recombinant P45 (rP45) possesses PEF activity comparable to a natural PEF. The results are shown in Figure 14. When no PEF activity is added, the 5.2 kb amplification product is not generated as shown in the lane labeled "none" of Figure 14. When proteins produced from the recombinant clones of P45 (called "1" or "23" from their plasmid delineation) were added to the On/Off assay in amounts of 5µl, 1 µl, or 1µl of a 1/10 dilution (approximately 100, 20 and 2 ng, respectively), they all showed enhancement of PCR product yield comparable to the two native PEF



-53-

preps. Interestingly, 5  $\mu$ l worked less well than 1  $\mu$ l. An "overdose" phenomenon was also noted with native PEF (example 16).

#### 4. Molecular Weight of Recombinant P45 Compared to Native PEF Complex and Protein Components of the Complex

The predicted molecular weight of the translated rP45 DNA sequence was 18.6 kDa. However, the native P45 component migrates as part of the PEF complex at 300 kDa without heat denaturation, and at 45 kDa with heat denaturation (99°C for 5 minutes, partially denatured). Protein complexes in hyperthermophiles are unusually stable and sometimes require extreme conditions before disassociation occurs. We have found that P45 migrates at approximately 18kD, or approximately 17-18kD, instead of 45kD when the native PEF complex is boiled in the presence of 2% SDS and 1% trichloroacetic acid (TCA) (fully denaturing conditions)(Figure 15). Without heat denaturation, more than half of the recombinant P45 migrated at approximately 50 kDa (45 + 4 kDa calmodulin binding tag), while the remaining sample migrated close to 26 kDa (Figure 15). With heat denaturation all of the sample migrated at approximately 26 kDa, suggesting that the recombinant is a dimer in its undenatured state. Trichloroacetic acid heat treatment failed to produce any additional bands in the recombinant sample.

The migration differences between the fully denatured forms of the native (18kD) and recombinant P45 (26kD) might be explained by differences in post-translational modifications between *P. furiosus* and the *E. coli* host or by the effects of the 4kD CBP (calmodulin binding protein) tag on the folding or migration of the rP45 fusion protein.

#### Example 11

##### Potential Mechanisms of PCR Enhancement by P45

##### 1. Identification of the Reaction Catalyzed by PEF/P45

The similarity of P45 to dCTP deaminases prompted us to test whether dCTP or other nucleotide triphosphates were substrates of PEF/P45. PEF was incubated with dCTP or dUTP and the reaction products were separated by reverse phase HPLC. Reaction mixtures (50  $\mu$ l) containing 1x cloned Pfu polymerase buffer, 10mM dCTP or dUTP, and 700ng PEF (or 5  $\mu$ l of cloned p45 (~100 ng) or nothing (neg. control)) were incubated at 72°C for 1 hour. The samples were injected on a 3.9 x 250 cm Waters Delta-pak C-18 column (300 angstrom/15  $\mu$ m), equilibrated in

-54-

50mM triethylammonium acetate, pH(7.0) (A). Products were eluted with acetonitrile (B) at 2 ml/min. using the following gradient: 0% B for 5 minutes, 0-10% B over 20 minutes. Absorbance of the eluate was monitored with a photodiode array detector, and peak areas were integrated at 260nm.

5           The data in Figure 16 show that both the native PEF complex and purified recombinant p45 utilize dUTP as a substrate. In the presence of cloned Pfu DNA polymerase PCR buffer, dUTP migrates at 10.517 minutes. After a one hour incubation with PEF or recombinant P45 (purified from clone 1 or 23), the dUTP peak disappeared (0.1% total peak area) and a new peak appeared eluting at  
10           4.400 minutes (98% total peak area). The product of the PEF + dUTP reaction migrated with the same retention time as a dUMP standard, which was different from the retention times of dUDP, dCMP, dCDP, and dCTP. To confirm that the product was dUMP, a dUMP standard and the PEF + dUTP product were mixed together and re-injected. The mixed sample produced only one peak.

15           In addition to dUTP, dCTP was also found to serve as a substrate for PEF (Figure 17). The product of the PEF + dCTP reaction appeared to be dCMP, based upon its retention time and absorbance maxima. Although PEF utilizes dCTP, dUTP is the preferred substrate of PEF/P45. In the example shown, only 16.6% of the dCTP was converted in 1 hour under the same conditions which converted  
20           100% dUTP (Figure 17, panel B). Moreover, when dCTP and dUTP were incubated together with PEF, only dUMP (4.2 minute peak in panel C) was generated. No reactivity with PEF was detected for the following nucleotides: dGTP, dATP, dTTP, dCMP, dUMP, and rUTP under the conditions used.

25           The substrate preference and reaction catalyzed by PEF/P45 was significantly different from that predicted based upon the amino acid similarity to dCTP deaminases. Although both enzymes bind dCTP and dUTP, the forward reaction catalyzed by dCTP deaminase is the deamination of dCTP to produce dUTP + NH<sub>3</sub>. PEF/P45, on the other hand, preferentially utilizes dUTP, and catalyzes the release of pyrophosphate (PPi). No deamination of dCTP by *P. furiosus* PEF or P45 was observed under the conditions used.  
30

-55-

## 2. Enzymatic Detection of Inorganic Pyrophosphate ( $PP_i$ ) Produced by PEF from dUTP

We tested whether inorganic pyrophosphate ( $PP_i$ ) or inorganic phosphate ( $P_i$ ) was generated during the PEF-catalyzed conversion of dUTP to dUMP (Figure 16). To assess whether  $P_i$  was produced, 900 or 1800 ng of PEF was incubated in the presence of 5 mM dUTP in 1x cloned Pfu polymerase buffer for 1 hour at 72°C. The reaction mixture was subsequently analyzed for the presence of  $P_i$  using the method of Heinonen and Lahti (Heinonen, J.K. and Lahti, R.J. (1981) *Anal. Biochem.* 113: 313-317). Previous results (HPLC) demonstrated that under the above reaction conditions, 100% of the dUTP should be converted to product. One hundred percent (100%) conversion corresponds to the production of 500 nmol of  $P_i$ , if  $P_i$  is a product of the reaction. However, no  $P_i$  was detected in these assays. The assay for  $P_i$  was sensitive enough to detect as little as about 50 nmoles of  $P_i$ . These observations show that  $P_i$  is not formed during the reaction of PEF with dUTP.

In experiments to investigate whether  $PP_i$  is formed during the reaction, 900 or 1800 ng of PEF was incubated with 10 mM dUTP in 1x cloned Pfu polymerase buffer for 1 hour at 72°C. The presence of  $PP_i$  in the reaction mixture was then quantified using Sigma's "Enzymatic Determination of Pyrophosphate" kit (Sigma Product No. P7275). The kit utilizes a coupled enzyme system whereby two moles of NADH are oxidized to NAD<sup>+</sup> for each mole of  $PP_i$  present in the reaction mixture. The oxidation of NADH is monitored spectrophotometrically at 340 nm. Using this assay system, the production of  $PP_i$  was clearly established in reactions that contained PEF and dUTP. No  $PP_i$  was detected in control reactions that lacked PEF or that contained PEF and dATP (in place of dUTP). The reactions that contained 1800 ng of PEF produced twice as much  $PP_i$  as those that contained 900 ng of PEF.

The temperature optimum ( $T_{opt}$ ) for the generation of  $PP_i$  from dUTP was measured by incubating mixtures of dUTP and native PEF at temperatures ranging from 73°C to 99°C.  $PP_i$  production by native PEF increased steadily over this temperature range and was highest at 99°C. While the  $T_{opt}$  for native PEF was found to >99°C, recombinant P45 exhibited maximal activity between 85°C and 93°C when tested at the same enzyme concentration (dUTP turn-over), but at a

-56-

lower total protein concentration. Accordingly, specific temperature ranges, such as from about 70°C to about 100°C, from about 85°C to about 93°C, or temperatures above about 70°C, can be used with a PEF or polymerase enhancing activity of the invention.

The enzyme activity of three different preparations of S200-purified native PEF was measured at 85°C. Protein concentrations were determined by both Bradford and by amino acid analysis. Shown below is a summary of the enzyme activity and specific activity (protein concentration determined by Bradford or AAA as indicated) of S200-purified native PEF. These activities were compared to the minimum amount of purified PEF required to amplify the 5.2kb target in the "on/off" assay described in example 1 (100µl PCR).

Prep	PPi production (µmole PPi/hr/µl)	Specific activity (µmole PPi/hr/µg) (Bradford)	Specific activity (µmole PPi/hr/µg) (AAA)	Minimum amount of dUTPase required for 5.2kb PCR (nmole PPi/hr)
1	1.23	1.76	4.03	0.11
2	0.27	0.59	3.13	0.11
3	0.14	na	1.37	0.22-0.44

The data indicate that for purified PEF preps 1 and 2, there is an excellent agreement between dUTPase activity and PCR enhancing activity. However, native PEF prep 3 exhibited 2-4 times less PCR enhancing activity, possibly due to its lower specific activity. Prep 3 may contain contaminants, which interfere with PCR enhancement.

PPi formation from dCTP was also measured by substituting dCTP for dUTP. For native PEF prep 1, the level of dCTPase activity was found to be 0.097(µmole PPi/hr/µl) at 85°C, which is 12-fold lower than the rate of PPi production from dUTP. In addition, recombinant P45 preparations were also tested and found to produce PPi from both dUTP and dCTP. Native PEF and recombinant P45 exhibited a similar degree of preference for dUTP, as compared to dCTP.

-57-

Therefore, recombinant P45 and structurally similar P45 proteins catalyze this reaction in the absence of any of the other components of the PEF complex. The methods described here for the PP<sub>i</sub> detection are, thus, a useful tool for analyzing the activity of or determining the presence of native PEF, PEF analog proteins, recombinant or synthetic PEF proteins, PEF complexes, and rP45.

### 3. PEF/dUTP Product Characterization with Mass Spectrometry

Electrospray mass spectral analysis was performed to characterize the byproduct of reactions employing PEF and dUTP. Analysis in the negative ionization mode produced a relatively clean spectrum exhibiting peaks at 307 and 615 m/e. These masses are consistent with those of dUMP [M-H]<sup>-</sup> and its non-covalent dimer [2M-H]<sup>-</sup>. Analysis in the positive mode gave a complex array of peaks.

### 4. Role of PEF/P45 in Eliminating dUTP Accumulation During PCR

An understanding of the catalytic activity of PEF/P45 has provided insight into the mechanism by which PCR enhancement occurs. Lasken *et al.* have reported that archeal DNA polymerases, such as Vent, incorporate dUTP at approximately 40% the rate of TTP (Roger S. Lasken, David M. Schuster, and Ayoub Rashtchian, (1996) *J. Biol. Chem.* 271; 17692-17696). However, further DNA synthesis by archeal DNA polymerases appears to be inhibited by dU-containing DNA. Inhibition appears related to the 6500-fold greater affinity of Vent for dU-containing sites, as compared to dT-containing DNA. Based upon these observations, Lasken has proposed that archeal DNA polymerases may play a role in repairing dU-containing DNA *in vivo*.

One possible mechanism for the PCR enhancing activity of PEF/P45 is that its associated dUTPase activity may convert any dUTP present during PCR to dUMP, a side-product which should not interfere with DNA polymerase activity. By so doing, dUTP would not be available for incorporation into the PCR product, and hence dU-DNA inhibition of the DNA polymerase would not occur. Such a mechanism is consistent with the increased PCR product yields generated by archeal DNA polymerase in the presence of PEF/P45.

dUTP, however, is not intentionally added to PCR reactions but may be generated by spontaneous deamination of dCTP during the high temperatures used in PCR cycling (Jens-Peter Horst and Hans-Joachim Fritz, (1996) *The EMBO*

-58-

*Journal* 15; 5459-5469). To assess the extent of dCTP deamination to dUTP during PCR, we incubated dCTP (10mM) at 95°C (PCR denaturation temperature) in the presence of cloned Pfu DNA polymerase PCR buffer. The products of heat treatment were analyzed by reverse phase HPLC (as described above) and peak assignments were made by comparing retention times to those of standards.

Figure 18 shows the results obtained when dCTP is heated for 1 hour (panel B) or 4 hours (panel C) at 95°C. dCTP breaks down into 2 predominate species, dCDP (22%/1 hr; 41%/4hr.) and dCMP (2.8%/1 hr.; 25%/4 hr.). A minor peak at 12.38 minutes was also produced (.049%/1 hr.; .116%/4 hr.), which was identified as dUTP, based upon its retention time (dUTP standard = 12.10 minutes; panel D) and absorbance maximum. As shown in Figure 19, the 12.38 minute peak generated by heating dCTP (spectrum 2) exhibits maximal absorbance at approximately 260nm, as does the dUTP standard (spectrum 3), while dCTP absorbs maximally at approximately 270nm (spectrum 1).

dUTP production was also measured during PCR cycling. A dCTP solution (10mM dCTP in cloned Pfu DNA polymerase PCR buffer) was subject to PCR cycling in a RoboCycler 40 using the cycling conditions described for the 5.2kb "on/off" system (example 1). Products generated during PCR were analyzed by HPLC as described above. After 30 cycles, the following products were evident: 79% dCTP, 19% dCDP, 1.8% dCMP, and .064% dUTP (Figure 20, panel A). The dUTP generated during PCR cycling of dCTP exhibited a retention time (12.1 min.) and absorbance maxima (263nm) which were very close to those of the dUTP standard (11.9 min.; 263nm) (Figure 20, panel D). When dCTP was cycled in the presence of Pfu DNA polymerase (50X PCR conditions, 10mM dCTP and 1.25U Pfu/ $\mu$ l reaction mix) there was no difference in the amount of dCDP, dCMP, or dUTP produced (Figure 20, panel C). However, when dCTP was cycled in the presence of PEF (50X PCR conditions, 10mM dCTP and 0.5ng PEF/ $\mu$ l reaction mix), the following products were produced: 73% dCTP, 19% dCDP, and 8% dCMP (Figure 20, panel B). There was no detectable dUTP generated when dCTP was cycled in the presence of PEF, consistent with an associated dUTPase activity of PEF.

The increase in dCMP production in dCTP + PEF samples (8%) as compared to dCTP  $\pm$  Pfu samples (1.7-1.8%) shows that in addition to eliminating

-59-

the minor dUTP deamination product during PCR, PEF will also convert dCTP to dCMP. In this experiment (50x PCR conditions), the final dCTP concentration post-PCR was 73% in PEF-containing reactions and 79% in those lacking PEF. This slight drop in the dCTP pool is not anticipated to affect PCR product yield or DNA polymerase replication fidelity significantly. However, it is anticipated that the use of higher amounts of PEF in PCR ( $>>1$  ng per 100  $\mu$ l reaction) will be deleterious due to dCTP reactivity. If high amounts of PEF are used, it is possible that the dCTP pool could fall below levels required for maximal yields and lowest misinsertion rates. As described elsewhere, we have observed inhibition of replication or amplification and/or smearing of products with the use of excessive amounts of PEF.

#### 5. PEF Reverses Inhibitory Action of dUTP in Amplification Reactions

Although Lasken reported that the incorporation of dUTP in the nascent DNA strand only inhibited archael polymerases by 40% (Lasken, et al. (1996) J. Biol. Chem. 271; 17692-17696), we unexpectedly found that the presence or addition of small amounts of dUTP into PCR reactions had more dramatic inhibitory consequences, as demonstrated below.

A relatively small (0.9kb) fragment of the human  $\alpha$ 1-antitrypsin gene was amplified in the absence or presence of dUTP. PCR reaction mixtures contained the following (in a 100  $\mu$ l volume): 1x Cloned Pfu polymerase buffer; 200  $\mu$ M each, dCTP, dGTP, dATP; 200 ng oligo F91-23 (100 ng/ $\mu$ l); 200 ng oligo R980-23 (100 ng/ $\mu$ l); 125 ng Human genomic DNA; 2.5 units Pfu DNA polymerase; 200  $\mu$ M total of (dTTP and dUTP) or (dTTP + PEF generated dUMP).

F91-23	5'	GAGGAGAGCAGGAAAGGTGGAAC	3'	(SEQ ID NO: 84)
R980-23	5'	CTCCATGTCCCAACTCCGATCAC	3'	(SEQ ID NO: 85)

PEF generated dUMP was prepared as described in Example 11, section 1, and purified by reverse phase HPLC.

PCR cycling was carried out as follows: 95°C for 1 minute (1 cycle); 95°C for 1 minute - 58°C for 1 minute - 72°C for 2 minutes (30 cycles).

The PCR products were examined on a 1% agarose, 1x TBE gel as shown in Figure 21. The amplification of the 900 bp product from human genomic DNA was completely inhibited with dUTP concentrations as low as 2  $\mu$ M (1% of the dTTP+dUTP pool) and partially inhibited at 0.2  $\mu$ M (0.1%). The concentration of

-60-

individual nucleotides in a standard PCR reaction is about 200µM. dUTP inhibition can be completely prevented when PEF or the recombinant P45 protein (rP45) is added to PCR reactions containing dUTP. Seven ng of PEF can reverse the inhibition caused by as much as 20 µM dUTP (data not shown). The preferred amount of PEF or P45 used in a particular reaction can be optimized according to the principles provided here or by methods for quantifying amplification reactions known in the art.

Unlike dUTP, the PEF generated byproduct, dUMP, was not inhibitory in Pfu polymerase-based PCR reactions, even when present at concentrations of 20 µM. In Figure 21, amplifications containing purified dUMP appear less robust than other bands on the gel. It should be noted that the other PCR products on this gel were generated by Pfu in the presence of PEF or rP45, which has been demonstrated to enhance PCR product yields as compared to amplifications conducted in the absence of PEF/rP45. These results are consistent with the enhancing activity of PEF and rP45 being a result of dUTPase activity. The dUTPase activity may hydrolyze dUTP and thereby prevent the incorporation of dUTP into DNA. As demonstrated in Figure 21, dUTP incorporation by Pfu DNA polymerase during PCR can significantly decrease PCR product yields. Accordingly, the invention comprises a method of enhancing nucleic acid replication or amplification reactions by reducing the dUTP concentration or preventing the incorporation of dUTP into replicated or amplified products, as well as compositions that are capable of preventing that incorporation.

We also tested whether PEF/rP45 could reverse the inhibition caused by uracil-containing DNA. PCR amplification was carried out in the presence of a third unrelated primer, which contains 9 dUs instead of dTs (dU oligo). Primers complementary to M13 DNA were synthesized.

Control Oligo 5' GGTTTTCCTCCAGTCACGACGTTGTAAAACGACGGCCAGT 3' (SEQ ID NO: 86)  
dU Oligo 5' GGUUUUCCAGUCACGACGUUGUAAAACGACGGCCAGU 3' (SEQ ID NO:87)

The 900 bp α1-antitrypsin fragment was amplified in the presence of the oligos, added at levels ranging from 200ng (16 pmole) to 0.2 ng (0.016 pmole) per 100ul reaction. In addition, similar reactions were performed with *Taq* DNA polymerase instead of *Pfu* DNA polymerase. In Figure 22, the control oligonucleotide shows inhibition of the Pfu-based PCR reaction at high



-61-

concentrations (16 and 8 pmole). In contrast, inhibition by the dU oligo was readily observed at concentrations as low as 0.8 pmole. It seems likely that both oligonucleotides sequester Pfu DNA polymerase away from the PCR primers and genomic DNA, but the dU oligo can inhibit at lower concentrations. These results show that Pfu DNA polymerase may be more tightly bound to a dU containing substrate. With the addition of PEF, a small amount of amplification product is observed in reactions containing 0.8 pmole of dU containing primer.

The enhancement by PEF in the dU oligo-inhibited reaction could be achieved through at least two possible pathways. The most likely explanation is that PEF is having no effect on the dU containing oligonucleotides and is simply increasing the activity of Pfu DNA polymerase by scavenging dUTP generated during PCR by heat- or chemically-induced deamination of dCTP (eg., lanes 2 and 3 of Figure 22). Or, PEF may work with Pfu DNA polymerase to remove the uracil moieties from the oligonucleotides through a repair pathway. Thus, PEF may also be used in repair reactions employing appropriate polymerase activities. Repair reactions are known in the art and methods to adopt the use of PEF into those reactions can be devised by those skilled in the art.

The knowledge of potential PEF mechanisms of action described here allows those skilled in the art to employ other dUTP modifying enzymes in enhancing polymerase reactions. A definition of one of these other modifying enzymes can be an enzyme that diminishes the capacity to incorporate dUTP by polymerases or at least partially inhibits dUTP incorporation. Assays used to identify and characterize PEF as described herein can also show other dUTP modifying enzymes. These other modifying enzymes could also mimic the enhancing attributes of PEF or a particular protein, such as P45 or rP45. An example of this class of enzyme would be dUTP pyrophosphatases (EC 3.6.1.23), such as deoxyuridine 5'-triphosphate nucleotide hydrolase, as well as other enzymes involved in dUTP metabolism, catabolism, or synthesis. These other enzymes may be used alone or in combination with PEF or other proteins or enhancing additives.

Furthermore, the presence of the consensus uridine-binding motif or the related sequences shown herein can also be used to define an enzyme or protein that is a PEF. Thus, a protein that comprises SEQ ID NO.: 72, or any one of SEQ

-62-

ID NOs.: 72-81, or combinations of these sequences, may be a PEF according to this invention.

### Example 12

#### Identification of Other Proteins for Enhancing Polymerase Activity

5       The structural information, in the amino acid and nucleotide sequences, as well as the functional information described here allow one skilled in the art to identify polymerase enhancing and/or dUTPase activities from a variety of sources. For example, we have shown above how degenerate probes made from the amino acid sequences of P50 and P45 can be used to clone nucleotide sequences  
10       encoding polymerase enhancing and dUTPase activities, or PEF. Since we have identified the importance of dUTPase activity in controlling and enhancing polymerase reactions, such as PCR, structural information available for any dUTPase can be put to a new and advantageous use in identifying and producing proteins for enhancing polymerization reactions. Furthermore, the assays  
15       described can be used to identify the presence of dUTPase activity from any source.

#### 1. Cloning Human dUTPase as a Representative Eukaryotic Protein for Enhancing Polymerization Reactions

20       To determine if other enzymes with dUTPase activity could also produce polymerase enhancing activity, we cloned a representative eukaryotic protein, human dUTPase. Total RNA was isolated from human placenta and converted to cDNA as follows: 5 µl total human RNA, 5 µl oligo dT (0.1 µg/µl), 1 µl Moloney murine leukemia virus reverse transcriptase (40 u/µl), 1 µl 100 mM dNTPs, 5 µl 10x first strand buffer, 33 µl DEPC-treated water (where 1x first strand buffer is 50 mM  
25       Tris-HCL (pH 8.3), 75 mM KCl, 10 mM DTT, and 3 mM MgCl<sub>2</sub>). The reaction was incubated at 37°C for one hour. A negative control was run in parallel without reverse transcriptase.

30       Primers containing a sequence specific to the 5' and 3' termini of one of the human dUTPase genes were synthesized and are shown below. The accession numbers for the cDNA sequence of Human deoxyuridine triphosphatase (DUT) are gi|1421817|gb|U62891|HSU62891. These primers also shared sequence with the vector pCAL-n-EK (in bold print below), which allowed ligation independent cloning (LIC) of the amplified product, as described in Example 10.

-63-

Primer 285 GACGACGACAAGATGCCCTGCTCTGAAGAGACACC (SEQ ID NO.: 88)  
Primer 286 GGAACAAGACCCGTTTAATTCTTTCCAGTGAACC (SEQ ID NO.: 89)

Prior to PCR, the reverse transcriptase was heat inactivated by incubating the reaction at 80°C for 5 minutes. The dUTPase sequence was amplified in a 100 µl reaction containing 1x cloned Pfu polymerase buffer, 200ng of each primer, 200 µM dNTPs, 2.5 units of Pfu DNA polymerase, 3 ng of PEF complex and 3µl of human placenta cDNA from the previous section.

The reactions were amplified under the following conditions: 95°C for 3 minutes (1 cycle); 95°C for 1 minute - 50°C for 1 minute - 72°C for 2 minutes (30 cycles). The amplified reaction was examined on a 1% agarose gel to confirm that the product exhibited the correct size before purification. The purified product was cloned into the expression vector pCAL-n-EK, as described in Example 10, and transformed into XL1-Blue cells. Three clones were confirmed to contain human dUTPase by sequencing of the first 500 bases. After the transformants were shown to contain the dUTPase sequence by PCR amplification, their plasmids were harvested and used to transform the *E. coli* strain BL21 (DE3).

## 2. Human dUTPase Expression and Activity Analysis

The BL21/dUTPase clones were induced with IPTG and the expressed protein was purified by means of the calmodulin binding peptide (CBP) tag expressed as a fusion protein at the amino terminus of the dUTPase sequence. The fusion protein was purified on calmodulin agarose, as described in example 10. The protein products were analyzed by SDS-PAGE and found to be of the correct molecular weight.

To confirm that the dUTPase clones were active, the Sigma pyrophosphatase assay (see Example 11) was utilized. The assay demonstrated that all of the clones tested could convert dUTP to dUMP + pyrophosphate. The human dUTPase enzyme was thermolabile and became completely inactive after a one minute pre-incubation at 70°C.

Polymerase enhancement was also detected with the 5.2 kb on/off assay. The assay was modified from the protocol described in Example 1 to allow detection of the thermolabile PEF activity. A PCR cocktail was mixed to provide an identical starting point for all samples. Ninety-nine microliters of the cocktail was aliquoted into six thin-walled, 0.5 ml tubes. The reactions contained 278 ng of

-64-

human genomic DNA, 200 ng of each primer (see Example 1), 200  $\mu$ M each dNTP, 2.5 units of Pfu DNA polymerase in 1x cloned Pfu polymerase buffer. At each 60°C annealing step, 0.5  $\mu$ l of one the following were added: human dUTPase preparation, a 1/10th dilution of the human dUTPase preparation, 2 ng/ $\mu$ l rP45 (positive control), or dUTPase storage buffer (negative control). Both human dUTPase reactions were run in duplicate. The samples were cycled as follows: 95°C for 1 minute (1 cycle); 95°C for 1 minute - 60°C for 1 minute - 72°C for 5.2 minutes (30 cycles).

10  $\mu$ l of each PCR reaction was visualized on a 1% agarose, 1x TBE gel by ethidium bromide staining. (See Figure 23) The undiluted human dUTPase preparation was able to enhance the polymerase activity to produce the 5.2 Kb band. As demonstrated by the negative control, lack of any enhancing factor results in a failed PCR. The positive control for this experiment, rP45, did not work in this experiment, presumably because an excessive quantity (30 ng total) was added.

### Example 13

#### Production of Antibodies to PEF and Western Blot Analysis

##### 1. Production of anti-PEF and anti-rP45 IgG

20 PEF-specific IgG was purified by immunoaffinity chromatography from the sera of rabbits previously immunized against a lot of native Pfu DNA polymerase containing PEF (see Figure 24 showing gel of purified native Pfu polymerase preps). The S-200-purified Pfu PEF was covalently coupled to AffiGel 10 (BioRad: Hercules, CA) in the presence of 20mM HEPES, 1mM DTT, 50mM KCl, 0.05% Tween 20, 1 mM EDTA, and 10% glycerol, following the manufacturer's recommended protocol. Rabbit sera (2.4 ml) was loaded onto a 0.2 ml column in the presence of 10mM Tris (pH 7.5). The column was washed extensively and the specific IgG was eluted with 0.1M glycine-HCl (pH 2.5) followed by 0.1M triethylamine (pH 11.5). Using a Centricon-30, the IgG was concentrated and the elution buffer replaced with PBS.

30 In addition, sera containing rP45-specific IgG was obtained by immunizing rabbits with recombinant P45, which was prepared as a tagged fusion protein, as described in example 10, section 2. The purified enzyme (0.177 mg/ml) was used to immunize two New Zealand white rabbits using the following immunization

-65-

schedule: 90 µg/rabbit in Complete Freund's Adjuvant (CFA); 18 days later, boost with 45 µg/rabbit in incomplete Freund's adjuvant (IFA); 39 days later, second boost; 45 days later, obtained serum sample for Western blot.

## 2. Western Blot Analysis Using anti-PEF Antibodies.

Cell extracts were prepared by suspending cells in 4X 50mM Tris, pH 8.2, 10mM BME, 1mM EDTA, and 10% glycerol, followed by sonication. Then, 2.5mM PMSF was added and the cellular debris removed by centrifugation for 15 minutes at 14,000 rpm. PEI was added to the supernatant to a final concentration of 0.9% and the mixture centrifuged again. The supernatants (10 µl) were electrophoresed on 4-20% SDS-PAGE gels and the proteins transferred to nitrocellulose by electroblotting. The blots were blocked with 1% Blotto/PBS for 1 hour at room temperature and then incubated with PEF-specific IgG overnight at 4°C. The blots were washed in PBS-.05% Tween 20, and then incubated with alkaline phosphatase-conjugated goat anti-rabbit IgG. The blot was washed and then incubated in color development solution (100 mM Tris-HCl, pH 9.5, 100 mM NaCl, 5 mM MgCl<sub>2</sub>, 0.3mg/ml NBT, and 0.15 mg/ml BCIP) for approximately 1-2 minutes. The enzyme reaction was stopped and the membrane was washed five times with deionized water.

Figure 25 depicts the results of the Western Blot. Extracts were prepared from 5 different fermentations of *P. furiosus* (lanes 1-5). In addition, three extracts or partially purified column fractions from *T. aquaticus* (lanes 6-8) and one extract from *E. coli* (lane 9) were also run. Purified PEF (550 ng) and pre-stained molecular weight markers were run in lanes 10 and 11, respectively. With the exception of the markers, all samples were boiled in SDS/BME dye prior to loading. The results show PEF-specific IgG binds to and cross-reacts with components of the PEF complex in crude Pfu extracts, including the P50 and P45 components. In contrast, no cross-reaction was observed with extracts from *T. aquaticus* or *E. coli*.

## 3. Western Blot Detection of Native PEF and Recombinant P45 Using anti-rP45 Sera

Native PEF samples were electrophoresed on a 4-20% gradient Tris-Glycine SDS gel. The samples were loaded without denaturation (P300 form) or after partial (boiling in 2%SDS; P45 form) or complete (boiling 2%SDS plus 1%TCA) denaturation. The samples were transferred to nitrocellulose and the blots

-66-

developed as described above, except that sera from rabbits immunized with recombinant P45 was used (diluted 1:1000).

In Figure 26, antibodies specific to recombinant P45 crossreact with bands in undenatured, partially denatured, and fully denatured samples of native PEF. The predominant bands are the P300 aggregate (P45/P50) present in the unheated lane, the 45kD (partially denatured) form present in the boiled/SDS sample, and the 17kD (fully denatured) form present in the boiled/1% TCA/2% SDS lane, which presumably represents the fully denatured monomer. In addition to these different aggregation states of native P45, there are also minor bands present in the Western blot, which may represent additional forms of P45 or the cross-reaction of antibodies to *Pyrococcus* proteins sharing common epitopes with *E. coli* contaminants present in the recombinant P45 preps.

As with the PEF-specific IgG from above, anti-rP45 sera can also be used to identify immunochemically-related proteins from other species. In Figure 27, the rP45-specific IgG was used to probe a blot containing cellular extracts from bacteria (*Bacillus* species, *T. aquaticus*), archaea (*P. furiosus*, *M. thermoautotrophicum*, and an uncharacterized species) and human cells (HL60 lymphoma). The results obtained show the antibodies bind to proteins from *P. furiosus*, the uncharacterized species, and the human cell line, which migrate with the same apparent molecular weight as purified native PEF. The antibodies also bind weakly with proteins in the *Bacillus* and *M. thermoautotrophicum* extracts, although these bands migrate with a different apparent molecular weight than *P. furiosus* PEF species. As with the PEF-specific IgG, no cross-reactivity was detected with *T. aquaticus* extracts.

In a separate Western assay, samples from *Thermus thermophilis* were run out on an SDS-PAGE gel and probed with the rP45 antisera. The rP45 antibody binds to a band of approximately 24kD (between markers of 16kD and 30kD) and another band approximately twice that size, presumably a dimeric or multimeric form. This demonstrates that PEF activity or protein is present in the *T. thermophilis* samples. The PEF as dUTPase activity in these samples was further confirmed by a dUTP conversion assay employing reverse phase HPLC. The assay detected the turn-over of dUTP with the *T. thermophilis* samples, but a control sample showed no detectable turn-over of the dUTP present.

-67-

**Example 14****Use of PEF Complex in Nucleic Acid Replication Reactions**

Initially and as a control to confirm the activity of the DNA polymerase used, gapped-duplex calf thymus DNA (Pharmacia) assays were performed. The polymerase cocktail contained 50mM Tris-HCl, pH 8.0, 5mM MgCl<sub>2</sub>, 1mM DTT, 50μg/ml BSA, 4% glycerol, 200μM each dNTP, [<sup>3</sup>H]TTP (.5mCi/μmole final concentration), and 250μg/ml of activated calf thymus DNA (Pharmacia). Samples containing Pfu DNA polymerase or *P. furiosus* PEF were serially diluted in Pfu DNA polymerase storage buffer (50mM Tris-HCl, pH 8.2, 0.1% NP-40, 0.1% Tween-20, 0.1mM EDTA, 1mM DTT, 50% glycerol) and then 1μl of each dilution was added to 10μl aliquots of polymerase cocktail. Polymerization reactions were conducted in triplicate for 30 minutes at 72°C. The extension reactions were quenched on ice, and then 5μl aliquots were spotted immediately onto DE81 filters (Whatman). Unincorporated [<sup>3</sup>H]TTP was removed by 6 washes with 2xSCC (0.3M NaCl, 30mM sodium citrate, pH 7.0), followed by one wash with 100% ethanol. Incorporated radioactivity was measured by scintillation counting. The assay was calibrated by counting a known amount of [<sup>3</sup>H]TTP on DE-81 filters, omitting the wash steps. One unit of polymerase activity is defined as the amount of enzyme which catalyzes the incorporation of 10 nmoles of total dNTP into polymeric form (binds to DE-81 paper) in 30 minutes at 72°C. Polymerase concentrations (U/ml) were extrapolated from the slope of the linear portion of units vs. enzyme volume plots.

The PEF samples tested exhibit no significant DNA polymerase activity while the Pfu DNA polymerase exhibited a specific activity of  $2-4 \times 10^4$  u/mg.

**1. Enhancement of Cloned Pfu DNA Polymerase with Pfu PEF**

*P. furiosus* PEF has been demonstrated to enhance the yield of PCR products generated with recombinant Pfu DNA polymerase using plasmid, lambda, and genomic DNA templates (Figures 28-30). The results demonstrate that the addition of *P. furiosus* PEF increases PCR product yield for a variety of PCR systems, ranging in target complexity. Relatively easy targets, e.g. plasmid DNA, can be successfully amplified with Pfu DNA polymerase, and the addition of PEF further increases product yield (Figure 28). We have found that fewer PCR cycles or lower template concentrations can be used in PEF-containing reactions,

-68-

demonstrating the advantageous properties of the proteins and compositions described and the methods employing them.

The most dramatic enhancements were observed when long (Figure 29) and/or highly complex targets (Figure 30) are amplified. The presence of PEF unexpectedly and significantly improves the amplification of these targets, which are often poorly amplified by single enzyme PCR reactions. In Figure 29, the addition of 1-100 ng of PEF (S200-purified PEF; prep. 3) to 100µl PCR reactions containing 5U of Pfu DNA polymerase significantly increased yields of a 10kb PCR product. In Figure 30, a 5.2kb target was successfully amplified from human genomic DNA in the presence of .3-280ng PEF (SCS #52 S200 purified) per 100µl PCR, but not in the absence of PEF, despite the use of 1.9 min. per kb extension times.

*P. furiosus* PEF has been found to enhance the yields of both specific and non-specific PCR products, when amplifications are conducted under less stringent PCR conditions with PCR primers that hybridize to more than one target. *E. coli* ssb (single-stranded binding protein; Stratagene's PerfectMatch) has been shown previously to increase the specificity of primer extension reactions, presumably by minimizing the formation of poorly matched primer-template complexes. When used in conjunction with *E. coli* ssb, *P. furiosus* PEF has been found to enhance the yield of specific PCR products (Figure 31). Pfu PEF also enhances yields of PCR products obtained with exo-Pfu and a mixtures of Taq and Pfu polymerase (for example, TaqPlus Long™, Stratagene; La Jolla, CA). Therefore, Pfu PEF is useful with polymerase mutants, truncated versions of polymerases, mixtures of polymerases, and polymerase-additive combinations (for example, Perfect Match®, Stratagene).

## 2. Enhancement of Native Pfu DNA Polymerase with Pfu PEF

Subsequent to identifying PEF from *P. furiosus*, we recognized that certain lots of native Pfu DNA polymerase preparations contained PEF. Varying amounts of the >250kD aggregate could be detected on silver-stained SDS-PAGE gels (e.g., lots 38 and 46 in Figure 24). Eleven of the 23 preparations examined were found to visibly contain low levels (8/11 lots; 0.1-1% total protein) to high levels (3/11 lots; 10-30% total protein) of PEF. PEF co-migrates with Pfu DNA polymerase during the initial Q- and SP-Sepharose columns in Pfu DNA polymerase purification, and



-69-

elutes just after the major peak of Pfu DNA polymerase activity on the Heparin Sepharose column (Figure 32). Pfu DNA polymerase fractions pooled after the Heparin Sepharose step are typically contaminated with varying amounts of PEF, depending upon the column efficiency and pooling strategy employed.

Contamination of native Pfu DNA polymerase with varying amounts of PEF could potentially contribute to lot-to-lot variation in the performance of native Pfu DNA polymerase in PCR. It is expected that lots containing approximately 1-100 ng of PEF per 2.5U of Pfu DNA polymerase will give rise to higher PCR product yields than amplifications conducted with cloned Pfu DNA polymerase or native Pfu DNA polymerase lots contaminated with  $\leq 10$  pg per 2.5U Pfu DNA polymerase ( $< .02$  % total protein). In theory, a lot containing certain PEF concentrations would exhibit reduced Pfu DNA polymerase performance, based upon the apparent inhibition of PEF at high concentrations discussed below ( $> 900$  ng per 2.5U Pfu DNA polymerase in 100 $\mu$ l PCRs).

When adding PEF to native Pfu DNA polymerase PCR amplifications, it is anticipated that the level of PEF contained in a particular lot of native Pfu must be taken into account to avoid smearing, inhibition of synthesis, or sub-optimal enhancement.

#### Example 15

##### Use of PEFs in Amplification Reactions

##### 1. Activity of Pfu PEF in a Standard PCR Protocol

To enhance PCR product yield, *P. furiosus* PEF is added, separately or pre-mixed with the DNA polymerase, to standard PCR amplifications. PCR amplification reactions generally consist of the following: .01-.05U/ $\mu$ l DNA polymerase, 2.5 pg/ $\mu$ l to 1 ng/ $\mu$ l *P. furiosus* PEF, 1-5 ng/ $\mu$ l of each primer, .5-10 ng/ $\mu$ l of genomic DNA template, and 0.2mM each dNTP in a suitable buffer (e.g., cloned Pfu DNA polymerase buffer consists of 10mM KCl, 10mM  $(\text{NH}_4)_2\text{SO}_4$ , 20mM Tris-HCl (pH 8.8), 2mM  $\text{MgSO}_4$ , 0.1%(v/v) Triton X-100, and 100ng/ $\mu$ l BSA). Amplifications from less-complex templates (e.g., lambda or plasmid DNA) are typically successful with 0.1-100 pg/ $\mu$ l DNA. The minimum amounts of native PEF typically used in PCR correspond to amounts of enzyme sufficient to catalyze the release of 0.1 to 0.4 nmole  $\text{PPi/hr}$  per 100 $\mu$ l PCR reaction (see example 11, section 2). PCR amplifications are conducted under standard conditions; e.g., 94-96°C for

-70-

0.5-2 min. (1 cycle)/ 94-96°C for 0.5-2 min.; 50-65°C for 0.5-2 min.; 68-72°C for 0.5-3 min. per kb of target amplified (30-40 cycles)/72°C for 0-10 min. (1 cycle).

## 2. Enhancement of PCR Amplification Reactions Employing DNA Polymerases Other than Pfu DNA Polymerase

5 *P. furiosus* PEF has been found to enhance the performance of other  $\alpha$ -type (Family B-related) DNA polymerases from thermophilic archaea. Enhanced PCR product yields were observed when *P. furiosus* PEF was added to amplifications conducted with DNA polymerases from both *Pyrococcus* and *Thermococcus* species. DNA polymerases demonstrated to function with *P. furiosus* PEF include: 10 Pwo DNA polymerase (Boehringer Mannheim; cloned from *P. woesei*), Deep Vent DNA polymerase (New England Biolabs; cloned from *P. sp. GBD*), JDF3 DNA polymerase (Stratagene; cloned from *P. sp. JDF3*), ES4 DNA polymerase (Stratagene; purified from *P. sp. ES4*, renamed *P. endeavori*) and Vent DNA polymerase (New England Biolabs; cloned from *T. litoralis*).

15 In Figure 33, S200-purified Pfu PEF (prep. 1) increased yields of a 6.2kb PCR product obtained with 2.5U of Pwo DNA polymerase. In Figure 34, the addition of S200-purified PEF (prep. 3) to JDF3 DNA polymerase PCRs increased the yield of a 10kb product amplified from lambda DNA (lanes 5-7; 11-13) and mouse genomic DNA containing 40 copies (lanes 8-10) or 1 copy (lanes 14-16) of a 20 lambda DNA transgene. In JDF3 DNA polymerase-based PCRs, amplifications are typically conducted with 1U of enzyme and extension times of 0.5 min./kb target. In Figure 35, the addition of 5 ng of S200-purified PEF (prep. 3) to ES4, JDF3, Pfu, and Vent DNA polymerase PCRs increased the yield of a 1.9kb product amplified from *P. furiosus* genomic DNA. In Figure 36, the addition of *P. furiosus* PEF was 25 also shown to increase yields of a 2kb PCR product amplified with Deep Vent DNA polymerase from transgenic mouse genomic DNA.

30 The addition of *P. furiosus* PEF may not enhance the yield of PCR products generated with Taq DNA polymerase (Figure 34, lanes 1-4 and Figure 23, lanes 9-10). Taq DNA polymerase is a Pol I-like (Family A-related) DNA polymerase isolated originally from the thermophilic eubacteria *Thermus aquaticus*.

## 3. Enhancement of RT-PCR Reactions

*P. furiosus* PEF has also been shown to enhance the yield of PCR products amplified from reverse transcribed RNA (cDNA) in a process known as RT-PCR,

-71-

known in the art. Enhancement has been observed in both 2-step (Figure 37) and 1-tube RT-PCR protocols (data not shown). In the former procedure, aliquots of cDNA synthesis reactions are added to PCR reactions containing a thermostable DNA polymerase (e.g., Pfu DNA polymerase) and *P. furiosus* PEF. In the latter approach, RNA is added to reaction mixtures containing a thermolabile RT, dNTPs, primers, a thermostable DNA polymerase (Pfu DNA polymerase), and *P. furiosus* PEF. cDNA synthesis and PCR amplification take place sequentially, in the same tube, by conducting cDNA synthesis at 37-42°C, followed by PCR amplification at elevated temperatures.

In the 2-step RT-PCR procedure, cDNA synthesis is first performed by combining the following reagents (50µl final volume): 5µg total RNA pre-annealed to 300 ng of primer (oligo dT, random hexamers, or a gene-specific primer), 4mM each dNTP, 20U RNase block (optional), and 50U MMLV RT (or other RT) in buffer containing 50mM Tris-HCl (pH 8.3), 75mM KCl, 3mM MgCl<sub>2</sub>, and DEPC-treated water. cDNA synthesis is allowed to proceed for 1 hour at 37-42°C. After heat inactivation of the RT, 1µl of cDNA is added to a 50µl PCR reaction containing 5U Pfu DNA polymerase, .01-50 ng *P. furiosus* PEF, 1µM of each primer, and 0.2mM each dNTP in buffer consisting of 10mM KCl, 10mM (NH<sub>4</sub>)<sub>2</sub>SO<sub>4</sub>, 20mM Tris-HCl (pH 8.8), 3mM MgSO<sub>4</sub>, 0.1%(v/v) Triton X-100, and 100µg/ml BSA. PCR can be conducted using the following cycle conditions: 94°C 2 min.; 60°C 2 min.; 68°C 1 min. per kb (1 cycle) and then 94°C 1 min.; 60°C 1min.; 68°C 1min. per kb (40 cycles).

The enhancement of RT-PCR with *P. furiosus* PEF was evaluated using PCR primers designed to span at least one intron-exon junction in the EGF receptor gene. Two primer sets were used (antisense: 5'GAG-TTA-AAT-GCC-TAC-ACT-GTA-TCT (SEQ ID NO.: 29); sense: 5'CAG-GAC-TCA-GAA-GCT-GCT-ATC-GAA (SEQ ID NO.: 30)(1kb) or 5'CTG-CAC-GTG-CCC-TGT-AGG-ATT-TGT (SEQ ID NO.: 31)(3kb)), which generate PCR products of 1kb or 3kb, as indicated, when amplification occurs from spliced RNA rather than contaminating DNA.

The PEF concentration which gives optimal performance was determined by titrating PEF preparation 3 (S-200 purified) and preparation 4 (heparin sepharose fraction) in the 2-step RT-PCR procedure described here. With PEF preparation 4, significant increase in the yield of the 1kb product was observed when 0.001-1µl

-72-

was added (10 pg-10 ng PEF) (Figure 37). Synthesis of the 3kb product was significantly enhanced when 0.1-1 $\mu$ l (1-10 ng PEF) of preparation 4 was added. With PEF preparation 3, significant increases in the yields of both the 0.6kb and the 3kb products were observed for all amounts tested in the range of 0.002-0.1 $\mu$ l (1-50 ng).

#### 4. Enhancement of Seamless™ Cloning Protocol

Seamless Cloning was performed using Stratagene's Seamless™ Cloning kit (Stratagene; La Jolla, CA, 1997/1998 Stratagene Catalog, specifically incorporated herein by reference), following the recommended protocol. The effect of *P. furiosus* PEF on the efficiency of Seamless Cloning is shown in Figure 38. Increased yield of a 7.2 kb PCR product was observed when 5 ng of S-200 purified PEF (prep. 1) was added to 50 $\mu$ l PCR reactions containing 2.5U Pfu DNA polymerase and methyl dCTP. Amplifications conducted in the presence of PEF utilized 1 min. per kb extension times. In the absence of PEF, very little PCR product was generated despite the use of longer 2 min./kb extension times.

#### 5. Enhancement of Linear Amplification Reactions: QuikChange™ Mutagenesis Protocol

Site-specific mutagenesis can be accomplished efficiently with double-stranded DNA templates using a linear amplification-based strategy employing Pfu DNA polymerase (QuikChange™ Site-Directed Mutagenesis Kit; Stratagene; La Jolla, CA, 1997/1998 Stratagene Catalog, specifically incorporated herein by reference). PCR primers containing the desired mutation(s) are designed to anneal to the same site on opposite strands. Primer extension reactions are conducted with a thermostable DNA polymerase (e.g. Pfu DNA polymerase) at temperatures which allow efficient synthesis in the absence of strand displacement activity (68°C). The amplification product is treated with DpnI to digest the parental methylated plasmid DNA and the resulting gapped, double-stranded DNA is then transformed into competent *E. coli* cells. Mutant clones are identified by DNA sequencing.

In evaluating *P. furiosus* PEF, mutagenesis was conducted using Stratagene's Quik Change mutagenesis kit, except that both recombinant and native Pfu DNA polymerase were used in the kit-supplied reaction buffer. The effect of *P. furiosus* PEF on the efficiency of QuikChange mutagenesis is shown in

-73-

Figure 39. The addition of .04 to 4 ng of PEF of PEF prep. 2 (heparin sepharose fraction lot 36 H.S. #78; 40ng/ $\mu$ l) to 50 $\mu$ l reactions increased the number of transformants generated by native and cloned Pfu DNA polymerases, while retaining mutation frequencies of 90-97%. Optimal results were obtained with 0.4 ng of PEF, which gave 7.5-fold and 5.3-fold increases in the number of mutant colonies generated with native and cloned Pfu DNA polymerase, respectively.

The use of PEFs in the QuikChange™ mutagenesis protocol corresponds to the use of PEFs in other linear amplification reactions known in the art, such as cycle sequencing reactions, primer extension reactions, and the like. PEFs can be employed in any linear amplification method to enhance the activity of the polymerase used. For example, the effect of *Pfu* PEF on cycle sequencing can be evaluated by comparing the quality and length of sequencing ladders generated with a polymerase, for example *exo*-*Pfu* DNA polymerase, in the absence and in the presence of PEF. A number of different cycle sequencing reactions, known to one skilled in the art, can be used in combination with the PEF complexes and proteins of this invention to enhance polymerase activity. In addition, primer extension reactions can also be enhanced with the use of PEFs. Numerous primer extension reactions are known in the art.

#### Example 16

##### Enhancing Titer of PEF

The nucleic acid replication enhancing activity of several different preparations of *Pfu* PEF have been evaluated in PCR, PCR-related applications, linear amplification-based applications, mutagenesis applications, cycle sequencing applications, and primer extension applications. One skilled in the art will appreciate that similar methods to optimize the use of any PEF, such as those specifically discussed herein, are apparent from the disclosure herein. A sample of substantially homogeneous PEF (e.g. S200-purified) enhances the performance of *Pfu* DNA polymerase in PCR amplification reactions when added at concentrations spanning a 10,000-fold range (0.09-900 ng/100 $\mu$ l). The highest yields of amplified product are observed in the presence of  $\approx$ 1 to 100 ng of P50. The addition of excess PEF ( $\geq$ 900 ng/100 $\mu$ l, where protein concentration was determined by the silver-staining intensity of the P50 band as compared to known protein standards)

-74-

or very low PEF concentrations ( $< 9 \text{ pg}/100\mu\text{l}$ ) in a PCR reaction resulted in lower PCR product yield.

The relative purity and PEF content of 4 preparations was examined by SDS-PAGE analysis (Figure 10). Preparations 1 and 3 consist of S200-purified PEF of  $>95\%$  homogeneity, while preparations 2 and 4 consist of concentrated heparin sepharose fractions of 10-20% homogeneity.

The PCR enhancing titer of S200-purified *P. furiosus* PEF (prep. 1; Figures 40 and 41) was determined using the F432-21/R6656-20/AAA742 primer-template system described in example 1. This preparation is approximately 95% pure, contains  $<.001\text{U}/\mu\text{l}$  DNA polymerase activity, and  $\approx 225 \text{ ng}/\mu\text{l}$  PEF. PCR enhancing activity was found to be optimal when  $.004\text{--}.4\mu\text{l}$  ( $.9\text{--}90 \text{ ng}$ ) of homogeneous *P. furiosus* PEF was added to  $100\mu\text{l}$  PCR reactions containing 2.5U of Pfu DNA polymerase. Reduced, but significant, PCR product yield was observed when  $4\mu\text{l}$  ( $900 \text{ ng}$ ) or  $.0004\mu\text{l}$  ( $.09 \text{ ng}$ ) of the S200-purified protein was added (Figure 40). In a second identical experiment, significant enhancement was noted when  $.004\text{--}.4\mu\text{l}$  was added, and very little improvement was noted with  $.0004\mu\text{l}$  (Figure 41).

The PCR enhancing titer of PEF preparation 2 was also determined (Figure 42-43). Preparation 2 consisted of a concentrated (Microcon-30) heparin sepharose fraction, in which *P. furiosus* PEF made up approximately 10% of the total protein and was present at approximately  $40 \text{ ng}/\mu\text{l}$ . Enhanced PCR product yield was observed when  $.0002\text{--}.2\mu\text{l}$  ( $.008\text{--}8 \text{ ng}$ ) of the column fraction was added to  $100\mu\text{l}$  PCR reactions containing 2.5U of Pfu DNA polymerase, with greatest enhancements observed in the range of  $.002\text{--}.2\mu\text{l}$  ( $.08\text{--}8 \text{ ng}$ ) (Figure 42). In a second identical experiment, enhanced PCR product yield was observed when  $.004\text{--}.04\mu\text{l}$  ( $.16\text{--}1.6 \text{ ng}$ ) of PEF preparation 2 was added to 2.5U of DNA polymerase per  $100\mu\text{l}$  PCR reaction (Figure 43). No PCR product was observed in the presence of  $4\mu\text{l}$  ( $160 \text{ ng}$ ) or  $.0004\mu\text{l}$  ( $.016 \text{ ng}$ ) of the column fraction, while a smear was generated when  $.4\mu\text{l}$  ( $16 \text{ ng}$ ) of the column fraction was added to PCR. Smear PCR products were also noted previously when  $1\mu\text{l}$  of heparin sepharose fractions containing the highest concentrations of PEF are added to PCRs (e.g., lot 37 H.S. fractions 69-81 in Figure 44; PEF prep. 2 in Figure 31).

In summary, substantially homogeneous *P. furiosus* PEF enhances the performance of Pfu DNA polymerase in test PCR amplifications when added at

-75-

concentrations spanning a 10,000-fold range (0.09-900 ng/100 $\mu$ l). The highest yields are observed in the presence of  $\approx$ 1 to 100 ng of PEF. The addition of excess PEF ( $\geq$ 900 ng/100 $\mu$ l) or PEF  $<$  9 pg/100 $\mu$ l PCR reaction, was found to give reduced performance (lower PCR product yield). Partially-purified PEF samples (heparin  
5 sepharose column fractions) also appear to enhance PCR product yield over a fairly broad range of PEF concentrations. With the column fraction analyzed here, highest yields of PCR were obtained in the range of .08 ng to 8 ng. The addition of higher amounts of the column fraction resulted in smearing (16-40 ng) or lack of enhancement (160 ng).

10 Inhibition of PCR enhancement at high concentrations of PEF appears to occur irrespective of the purity of the PEF sample. The addition of higher concentrations of homogeneous PEF ( $\geq$ 900 ng) resulted in lower yields of PCR product than could be obtained with  $<$ 900 ng PEF. Heparin sepharose fractions of 10-20% purity also gave reduced PCR product yields when high amounts of PEF  
15 were added. Up to 8 ng of PEF in prep. 2 (H.S. #78 fraction) could be added before smearing or inhibition occurred. The discrepancy between the amount of PEF which is inhibitory in homogeneous preparations ( $\geq$ 900 ng), as compared to partially-purified column fractions ( $>$ 16 ng), suggests that additional protein or DNA contaminants may be present in the heparin sepharose fractions.

20 Examination of heparin sepharose fractions revealed that *P. furiosus* PEF elutes just after the major peak of Pfu DNA polymerase activity (e.g., fractions 50-61 in Figure 32). SDS-PAGE analysis showed that the highest levels of the  $>$ 250kD PEF appeared in fractions 63-78 from the native Pfu DNA polymerase purification (Figure 32). PCR enhancing activity was observed in fractions 37-90 (Figure 44).  
25 Discrete PCR product bands were generated with fractions 37-48 and 87-90, which contain very low levels of PEF. Interestingly, DNA smears were generated with fractions 69-81 (Figure 44), which contain the highest levels of PEF and no detectable Pfu DNA polymerase. These results are consistent with the hypothesis that *P. furiosus* PEF acts as an inhibitory substance when present at high  
30 concentrations. However, concentrations high enough to adversely effect polymerization and PCR reactions can easily be determined through the methods and assays described herein. Thus, one skilled in the art can avoid concentrations

-76-

of PEF that inhibit polymerase activity or PCR reactions without undue experimentation.

As observed with PCR, inhibition during linear amplification protocols was noted with high concentrations of PEF-containing heparin sepharose fractions (Figure 39). The addition of 40 ng (1 $\mu$ l prep. #2) of PEF to QuickChange™ reactions resulted in reduced yield of amplification product, as visualized by a reduction in the ethidium bromide-staining intensity of DNA bands on agarose gels. Reduced yield accompanied a 1.8 to 2.8-fold reduction in the number of transformants and a slight, but reproducible, decrease in mutation frequency.

A possible and the most likely explanation for inhibition by homogenous PEF preparations is depletion of dCTP. In Example 11, section 1, we demonstrated that PEF can utilize dCTP as a substrate, although much less efficiently than dUTP. At high PEF concentrations it is possible that enough dCTP is hydrolyzed by PEF to drop the dCTP levels below what is required for optimal DNA synthesis. It is also possible that moderate to high levels of PEF could alter dCTP levels enough to affect DNA polymerase misincorporation rates. Alternatively, contaminants in the substantially homogenous PEF preparations may also cause the inhibition and may only be present in sufficient concentrations when high concentrations of PEF are used.

Each of the references referred to herein can be relied on by one skilled in the art in making and using embodiments of the invention. In addition, each reference is specifically incorporated, in its entirety, into this disclosure.

The sequence listing information that follows incorporates the sequences in prior U.S. Patent application serial no. 08/822,744, which is specifically incorporated herein by reference. The sequence information from any one sequence or any combination of sequences can be translated into a computer readable medium by those of skill in the art. Furthermore, the sequences of the specific clones or plasmids described or identified herein can be easily determined and used in a computer readable medium by one skilled in the art.



-77-

## SEQUENCE LISTING

## (1) GENERAL INFORMATION:

- (i) APPLICANT(S): HOGREFE, Holly  
HANSEN, Connie J
- (ii) TITLE OF INVENTION: Polymerase Enhancing Factor (PEF)  
Extracts, PEF Protein Complexes,  
Isolated PEF Proteins, and Methods for  
Purifying and Identifying Them

(iii) NUMBER OF SEQUENCES: 89

## (iv) CORRESPONDENCE ADDRESS:

- (A) ADDRESSEE: M. Paul Barker, Esq., Finnegan,  
Henderson, Farabow, Garrett & Dunner, LLP
- (B) STREET: 1300 I. Street, N.W.
- (C) CITY: Washington
- (D) STATE: DC
- (F) ZIP: 20005

## (v) COMPUTER READABLE FORM:

- (A) MEDIUM TYPE: Floppy disk
- (B) COMPUTER: IBM PC compatible
- (C) OPERATING SYSTEM: PC-DOS/MS-DOS
- (D) SOFTWARE: PatentIn Release #1.0, Version #1.25

## (vi) CURRENT APPLICATION DATA:

- (A) APPLICATION NUMBER: NOT YET KNOWN
- (B) FILING DATE: 20-MARCH-1998
- (C) CLASSIFICATION:

## (viii) ATTORNEY/AGENT INFORMATION:

- (A) NAME: BARKER, M. PAUL
- (B) REGISTRATION NUMBER: 32,013
- (C) REFERENCE/DOCKET NUMBER: 4121.0116.02304

## (ix) TELECOMMUNICATION INFORMATION:

- (A) TELEPHONE: 202 408-4000
- (B) TELEFAX: 202 408-4400

## (2) INFORMATION FOR SEQ ID NO:1:

- (i) SEQUENCE CHARACTERISTICS:
- (A) LENGTH: 15 amino acids
- (B) TYPE: amino acid
- (C) STRANDEDNESS: unknown
- (D) TOPOLOGY: unknown

(ii) MOLECULE TYPE: peptide

(iii) HYPOTHETICAL: NO

(iv) ANTI-SENSE: NO

(v) FRAGMENT TYPE: N-terminal

## (xi) SEQUENCE DESCRIPTION: SEQ ID NO:1:

Xaa Xaa Leu His His Val Lys Leu Ile Tyr Ala Thr Xaa Xaa Xaa  
1 5 10 15

## (2) INFORMATION FOR SEQ ID NO:2:

- (i) SEQUENCE CHARACTERISTICS:
- (A) LENGTH: 15 amino acids
- (B) TYPE: amino acid
- (D) TOPOLOGY: unknown

(ii) MOLECULE TYPE: peptide

(iii) HYPOTHETICAL: NO

(iv) ANTI-SENSE: NO

(v) FRAGMENT TYPE: N-terminal

-78-

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:2:

Xaa Xaa Xaa Pro Asp Trp Xaa Xaa Arg Xaa Glu Xaa Leu Xaa Xaa  
 1 5 10 15

(2) INFORMATION FOR SEQ ID NO:3:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 35 amino acids  
 (B) TYPE: amino acid  
 (D) TOPOLOGY: unknown

(ii) MOLECULE TYPE: peptide

(iii) HYPOTHETICAL: NO

(iv) ANTI-SENSE: NO

(v) FRAGMENT TYPE: N-terminal

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:3:

Xaa Leu Leu His His Val Lys Leu Ile Tyr Ala Thr Lys Xaa Arg Xaa  
 1 5 10 15  
 Leu Val Gly Lys Xaa Ile Val Leu Ala Ile Pro Gly Xaa Xaa Ala Xaa  
 20 25 30

Xaa Xaa Xaa

35

(2) INFORMATION FOR SEQ ID NO:4:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 18 amino acids  
 (B) TYPE: amino acid  
 (D) TOPOLOGY: unknown

(ii) MOLECULE TYPE: peptide

(iii) HYPOTHETICAL: NO

(iv) ANTI-SENSE: NO

(v) FRAGMENT TYPE: N-terminal

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:4:

Xaa Xaa Xaa Pro Asp Trp Xaa Xaa Arg Xaa Glu Xaa Leu Xaa Glu Xaa  
 1 5 10 15

Xaa Xaa

(2) INFORMATION FOR SEQ ID NO:5:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 17 amino acids  
 (B) TYPE: amino acid  
 (D) TOPOLOGY: unknown

(ii) MOLECULE TYPE: peptide

(iii) HYPOTHETICAL: NO

(iv) ANTI-SENSE: NO

(v) FRAGMENT TYPE: internal

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:5:

Xaa Tyr Asp Ala Val Ile Met Ala Ala Ala Val Val Asp Phe Arg Pro  
 1 5 10 15

Lys

(2) INFORMATION FOR SEQ ID NO:6:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 24 amino acids  
 (B) TYPE: amino acid  
 (D) TOPOLOGY: unknown

(ii) MOLECULE TYPE: peptide

(iii) HYPOTHETICAL: NO

(iv) ANTI-SENSE: NO

(v) FRAGMENT TYPE: internal

-79-

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:6:

Ala Asp Leu Val Val Gly Asn Thr Leu Glu Ala Phe Gly Ser Glu Glu  
 1 5 10 15

Asn Gln Val Val Leu Ile Gly Arg  
 20

(2) INFORMATION FOR SEQ ID NO:7:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 17 amino acids  
 (B) TYPE: amino acid  
 (D) TOPOLOGY: unknown

(ii) MOLECULE TYPE: peptide

(iii) HYPOTHETICAL: NO

(iv) ANTI-SENSE: NO

(v) FRAGMENT TYPE: N-terminal

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:7:

Gly Ala Met Leu His His Val Lys Leu Ile Tyr Ala Xaa Lys Leu Arg  
 1 5 10 15

Lys

(2) INFORMATION FOR SEQ ID NO:8:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 18 amino acids  
 (B) TYPE: amino acid  
 (D) TOPOLOGY: unknown

(ii) MOLECULE TYPE: peptide

(iii) HYPOTHETICAL: NO

(iv) ANTI-SENSE: NO

(v) FRAGMENT TYPE: N-terminal

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:8:

Gly Ala Met Leu His His Val Lys Leu Ile Tyr Ala Thr Lys Xaa Xaa  
 1 5 10 15

Arg Lys

(2) INFORMATION FOR SEQ ID NO:9:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 13 amino acids  
 (B) TYPE: amino acid  
 (D) TOPOLOGY: unknown

(ii) MOLECULE TYPE: peptide

(iii) HYPOTHETICAL: NO

(iv) ANTI-SENSE: NO

(v) FRAGMENT TYPE: N-terminal

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:9:

Met Leu His His Val Lys Leu Ile Tyr Ala Thr Lys Leu  
 1 5 10

(2) INFORMATION FOR SEQ ID NO:10:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 16 amino acids  
 (B) TYPE: amino acid  
 (D) TOPOLOGY: unknown

(ii) MOLECULE TYPE: peptide

(iii) HYPOTHETICAL: NO

(iv) ANTI-SENSE: NO

(v) FRAGMENT TYPE: N-terminal

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:10:

Gly Xaa Xaa Xaa Pro Asp Trp Xaa Xaa Lys Phe Arg Lys Glu Glu Ser

-80-

- 1 5 10 15
- (2) INFORMATION FOR SEQ ID NO:11:
- (i) SEQUENCE CHARACTERISTICS:
- (A) LENGTH: 17 amino acids
- (B) TYPE: amino acid
- (D) TOPOLOGY: unknown
- (ii) MOLECULE TYPE: peptide
- (iii) HYPOTHETICAL: NO
- (iv) ANTI-SENSE: NO
- (v) FRAGMENT TYPE: N-terminal
- (xi) SEQUENCE DESCRIPTION: SEQ ID NO:11:
- Gly Ala Ile Leu Leu Pro Asp Trp Lys Ile Arg Lys Glu Ile Leu Ile
- 1 5 10 15
- Glu
- 15 (2) INFORMATION FOR SEQ ID NO:12:
- (i) SEQUENCE CHARACTERISTICS:
- (A) LENGTH: 16 amino acids
- (B) TYPE: amino acid
- (D) TOPOLOGY: unknown
- (ii) MOLECULE TYPE: peptide
- (iii) HYPOTHETICAL: NO
- (iv) ANTI-SENSE: NO
- (v) FRAGMENT TYPE: N-terminal
- (xi) SEQUENCE DESCRIPTION: SEQ ID NO:12:
- Xaa Met His His Ile Lys Leu Xaa Tyr Ala Thr Xaa Ser Arg Lys
- 1 5 10 15
- (2) INFORMATION FOR SEQ ID NO:13:
- (i) SEQUENCE CHARACTERISTICS:
- (A) LENGTH: 18 amino acids
- (B) TYPE: amino acid
- (D) TOPOLOGY: unknown
- (ii) MOLECULE TYPE: peptide
- (iii) HYPOTHETICAL: NO
- (iv) ANTI-SENSE: NO
- (v) FRAGMENT TYPE: N-terminal
- (xi) SEQUENCE DESCRIPTION: SEQ ID NO:13:
- Met Leu Tyr Leu Val Arg Pro Asp Trp Lys Arg Arg Lys Glu Ile Leu
- 1 5 10 15
- Ile Glu
- 40 (2) INFORMATION FOR SEQ ID NO:14:
- (i) SEQUENCE CHARACTERISTICS:
- (A) LENGTH: 23 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: unknown
- (ii) MOLECULE TYPE: DNA (genomic)
- (iii) HYPOTHETICAL: NO
- (iv) ANTI-SENSE: NO
- (xi) SEQUENCE DESCRIPTION: SEQ ID NO:14:
- CAYCAYGAHA ARYTHATTTA CGC
- 50 (2) INFORMATION FOR SEQ ID NO:15:
- (i) SEQUENCE CHARACTERISTICS:
- (A) LENGTH: 23 base pairs
- (B) TYPE: nucleic acid

23

-81-

(C) STRANDEDNESS: single  
 (D) TOPOLOGY: unknown  
 (ii) MOLECULE TYPE: DNA (genomic)  
 (iii) HYPOTHETICAL: NO  
 (iv) ANTI-SENSE: YES  
 (xi) SEQUENCE DESCRIPTION: SEQ ID NO:15:  
 GCCATDATNA CDGCRTCGTA TTT 23

(2) INFORMATION FOR SEQ ID NO:16:  
 (i) SEQUENCE CHARACTERISTICS:  
 (A) LENGTH: 23 base pairs  
 (B) TYPE: nucleic acid  
 (C) STRANDEDNESS: single  
 (D) TOPOLOGY: unknown  
 (ii) MOLECULE TYPE: DNA (genomic)  
 (iii) HYPOTHETICAL: NO  
 (iv) ANTI-SENSE: NO  
 (xi) SEQUENCE DESCRIPTION: SEQ ID NO:16:  
 CAYCAYGAHA ARYTHATATA CGC 23

(2) INFORMATION FOR SEQ ID NO:17:  
 (i) SEQUENCE CHARACTERISTICS:  
 (A) LENGTH: 20 base pairs  
 (B) TYPE: nucleic acid  
 (C) STRANDEDNESS: single  
 (D) TOPOLOGY: unknown  
 (ii) MOLECULE TYPE: DNA (genomic)  
 (iii) HYPOTHETICAL: NO  
 (iv) ANTI-SENSE: YES  
 (xi) SEQUENCE DESCRIPTION: SEQ ID NO:17:  
 ARDACDACYT GRTTTTCTTC 20

(2) INFORMATION FOR SEQ ID NO:18:  
 (i) SEQUENCE CHARACTERISTICS:  
 (A) LENGTH: 1209 base pairs  
 (B) TYPE: nucleic acid  
 (C) STRANDEDNESS: single  
 (D) TOPOLOGY: unknown  
 (ii) MOLECULE TYPE: DNA (genomic)  
 (iii) HYPOTHETICAL: NO  
 (iv) ANTI-SENSE: NO  
 (xi) SEQUENCE DESCRIPTION: SEQ ID NO:18:

ATGCTTCACC	ACGTCAAGCT	AATCTACGCC	ACAAAAAGTC	GAAAGCTAGT	TGGAAAAAAG	60
ATAGTCNNNN	NNNNNCCAGG	GAGTATTGCG	GCTTTGGATG	TGAAAGCTTG	TGAGGGGACTA	120
ATTAGGCATG	GGGCCGAAGT	TCATGCAGTG	ATGAGTGAGG	CAGCCACCAA	GATAATTCAT	180
CCTTATGCAT	GGAATTTGCC	CACGGGAAAT	CCAGTCATAA	CTGAGATCAC	TGGATTTATC	240
GAGCATGTTG	AGTTAGCAGG	GGAACATGAG	AATAAAGCAG	ATTTAATTTT	GGTTTGTCTT	300
GCCACTGCCA	ACACAATTAG	TAAGATTGCA	TGTGGAATAG	ATGATACTCC	AGTAACTACA	360
GTCGTGACCA	CAGCATTTC	CCACATTCCA	ATTATGATAG	CCCCAGCAAT	GCATGAGACA	420
ATGTACAGGC	ATCCCATAGT	AAGGGAGAAC	ATTGAAAGGT	TAAAGAAGCT	TGGCGTTGAG	480
TTTATAGGAC	CAAGAATTGA	GGAGGGAAAG	GCAAAAGTTG	CAAGCATTGA	TGAAATAGTT	540
TACAGAGTTA	TTAAAAAGCT	CCACAAAAAA	ACATTGGAAG	GGAAGAGAGT	CCTAGTAACG	600
GCGGGAGCAA	CAAGAGAGTA	CATAGATCCA	ATAAGATTCA	TAACAAATGC	CAGCAGTGGA	660
AAAATGGGAG	TAGCGTTGGC	TGAAGAAGCA	GATTTTAGAG	GAGCTGTTAC	CCTCATAAGA	720
ACAAAGGGAA	GTGTAAAGGC	TTTTAGAATC	AGAAAAATCA	AATTGAAGGT	TGAGACAGTG	780
GAAGAAATGC	TTTCAGCGAT	TGAAAATGAG	TTGAGGAGTA	AAAAGTATGA	CGTAGTTATT	840
ATGGCAGCTG	CTGTAAGCGA	TTTTAGGCCA	AAAATTAAAG	CAGAGGGAAA	AATTAAAAGC	900

-82-

GGAAGATCAA TAACGATAGA GCTCGTTCCN NNNAAATCCCA AAATCATTGA TAGAATAAAG 960  
 GAAATTCAAC CAAATGTCTT TCTTGTGTTGA TTAAAGCAG AAACCTTCAA AGAAAAGCTT 1020  
 ATAGAAGAAG GTAAAAGGCA GATTGAGAGG GCCAAGGCTG ACTTAGTCGT TGGTAACACA 1080  
 TTGGAAGCCT TTGGAAGCGA GGAAAACCAA GTAGTATTAA TTGGCAGAGA TTTCACAAAA 1140  
 GAACTTCCAA AAATGAAAAA GAGAGAGTTA GCAGAGAGAA TTTGGGATGA GATAGAGAAA 1200  
 TTNCTGTCC 1209

(2) INFORMATION FOR SEQ ID NO:19:

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 403 amino acids

(B) TYPE: amino acid

(D) TOPOLOGY: unknown

(ii) MOLECULE TYPE: protein

(iii) HYPOTHETICAL: NO

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:19:

Met	Leu	His	His	Val	Lys	Leu	Ile	Tyr	Ala	Thr	Lys	Ser	Arg	Lys	Leu
1				5					10					15	
Val	Gly	Lys	Lys	Ile	Val	Xaa	Xaa	Xaa	Pro	Gly	Ser	Ile	Ala	Ala	Leu
				20				25					30		
Asp	Val	Lys	Ala	Cys	Glu	Gly	Leu	Ile	Arg	His	Gly	Ala	Glu	Val	His
				35			40					45			
Ala	Val	Met	Ser	Glu	Ala	Ala	Thr	Lys	Ile	Ile	His	Pro	Tyr	Ala	Trp
		50				55					60				
Asn	Leu	Pro	Thr	Gly	Asn	Pro	Val	Ile	Thr	Glu	Ile	Thr	Gly	Phe	Ile
65				70						75				80	
Glu	His	Val	Glu	Leu	Ala	Gly	Glu	His	Glu	Asn	Lys	Ala	Asp	Leu	Ile
				85					90					95	
Leu	Val	Cys	Pro	Ala	Thr	Ala	Asn	Thr	Ile	Ser	Lys	Ile	Ala	Cys	Gly
			100					105					110		
Ile	Asp	Asp	Thr	Pro	Val	Thr	Thr	Val	Val	Thr	Thr	Ala	Phe	Pro	His
			115				120					125			
Ile	Pro	Ile	Met	Ile	Ala	Pro	Ala	Met	His	Glu	Thr	Met	Tyr	Arg	His
			130			135					140				
Pro	Ile	Val	Arg	Glu	Asn	Ile	Glu	Arg	Leu	Lys	Lys	Leu	Gly	Val	Glu
				145		150			155				160		
Phe	Ile	Gly	Pro	Arg	Ile	Glu	Glu	Gly	Arg	Ala	Lys	Val	Ala	Ser	Ile
				165				170					175		
Asp	Glu	Ile	Val	Tyr	Arg	Val	Ile	Lys	Lys	Leu	His	Lys	Lys	Thr	Leu
			180				185					190			
Glu	Gly	Lys	Arg	Val	Leu	Val	Thr	Ala	Gly	Ala	Thr	Arg	Glu	Tyr	Ile
			195				200					205			
Asp	Pro	Ile	Arg	Phe	Ile	Thr	Asn	Ala	Ser	Ser	Gly	Lys	Met	Gly	Val
				210		215					220				
Ala	Leu	Ala	Glu	Glu	Ala	Asp	Phe	Arg	Gly	Ala	Val	Thr	Leu	Ile	Arg
				225		230			235				240		
Thr	Lys	Gly	Ser	Val	Lys	Ala	Phe	Arg	Ile	Arg	Lys	Ile	Lys	Leu	Lys
				245				250					255		
Val	Glu	Thr	Val	Glu	Glu	Met	Leu	Ser	Ala	Ile	Glu	Asn	Glu	Leu	Arg
			260			265						270			
Ser	Lys	Lys	Tyr	Asp	Val	Val	Ile	Met	Ala	Ala	Ala	Val	Ser	Asp	Phe
			275			280						285			
Arg	Pro	Lys	Ile	Lys	Ala	Glu	Gly	Lys	Ile	Lys	Ser	Gly	Arg	Ser	Ile
			290			295					300				
Thr	Ile	Glu	Leu	Val	Pro	Xaa	Asn	Pro	Lys	Ile	Ile	Asp	Arg	Ile	Lys

-83-

305                      310                      315                      320  
 Glu Ile Gln Pro Asn Val Phe Leu Val Gly Phe Lys Ala Glu Thr Ser  
                                  325                      330                      335  
 Lys Glu Lys Leu Ile Glu Glu Gly Lys Arg Gln Ile Glu Arg Ala Lys  
                                  340                      345                      350  
 Ala Asp Leu Val Val Gly Asn Thr Leu Glu Ala Phe Gly Ser Glu Glu  
                                  355                      360                      365

Asn Gln Val Val Leu Ile Gly Arg Asp Phe Thr Lys Glu Leu Pro Lys  
                                  370                      375                      380  
 Met Lys Lys Arg Glu Leu Ala Glu Arg Ile Trp Asp Glu Ile Glu Lys  
 385                      390                      395                      400  
 Xaa Leu Ser

## (2) INFORMATION FOR SEQ ID NO:20:

## (i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 33 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: unknown

(ii) MOLECULE TYPE: DNA (genomic)

(iii) HYPOTHETICAL: NO

(iv) ANTI-SENSE: NO

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:20:

CATAGCGAAT TCGCAAAC TTTGCGGTA TGG

33

## (2) INFORMATION FOR SEQ ID NO:21:

## (i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 33 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: unknown

(ii) MOLECULE TYPE: DNA (genomic)

(iii) HYPOTHETICAL: NO

(iv) ANTI-SENSE: YES

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:21:

ACTACGGAAT TCCACGAAA ATGCCGCTCA TCC

33

## (2) INFORMATION FOR SEQ ID NO:22:

## (i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 20 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: unknown

(ii) MOLECULE TYPE: DNA (genomic)

(iii) HYPOTHETICAL: NO

(iv) ANTI-SENSE: NO

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:22:

GGCGTTTCCG TTCTTCTTCG

20

## (2) INFORMATION FOR SEQ ID NO:23:

## (i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 20 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: unknown

(ii) MOLECULE TYPE: DNA (genomic)

(iii) HYPOTHETICAL: NO

-84-

(iv) ANTI-SENSE: YES

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:23:

CCATCTCACG CGCCAGTTTC

20

(2) INFORMATION FOR SEQ ID NO:24:

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 23 base pairs

(B) TYPE: nucleic acid

(C) STRANDEDNESS: single

(D) TOPOLOGY: unknown

(ii) MOLECULE TYPE: DNA (genomic)

(iii) HYPOTHETICAL: NO

(iv) ANTI-SENSE: NO

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:24:

GAGGAGAGCA GGAAAGGTGG AAC

23

(2) INFORMATION FOR SEQ ID NO:25:

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 21 base pairs

(B) TYPE: nucleic acid

(C) STRANDEDNESS: single

(D) TOPOLOGY: unknown

(ii) MOLECULE TYPE: DNA (genomic)

(iii) HYPOTHETICAL: NO

(iv) ANTI-SENSE: YES

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:25:

GCTGGGAGAA GACTTCACTG G

21

(2) INFORMATION FOR SEQ ID NO:26:

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 19 base pairs

(B) TYPE: nucleic acid

(C) STRANDEDNESS: single

(D) TOPOLOGY: unknown

(ii) MOLECULE TYPE: DNA (genomic)

(iii) HYPOTHETICAL: NO

(iv) ANTI-SENSE: NO

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:26:

GAGCTTGCTC AACTTTATC

19

(2) INFORMATION FOR SEQ ID NO:27:

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 23 base pairs

(B) TYPE: nucleic acid

(C) STRANDEDNESS: single

(D) TOPOLOGY: unknown

(ii) MOLECULE TYPE: DNA (genomic)

(iii) HYPOTHETICAL: NO

(iv) ANTI-SENSE: YES

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:27:

GATAGAGATA GTTCTGAG ACG

23

(2) INFORMATION FOR SEQ ID NO:28:

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 23 base pairs

(B) TYPE: nucleic acid

(C) STRANDEDNESS: single

(D) TOPOLOGY: unknown

(ii) MOLECULE TYPE: DNA (genomic)



-85-

(iii) HYPOTHETICAL: NO  
(iv) ANTI-SENSE: YES  
(xi) SEQUENCE DESCRIPTION: SEQ ID NO:28:  
CGGGATATCG ACATTCTGC ACC 23

5 (2) INFORMATION FOR SEQ ID NO:29:  
    (i) SEQUENCE CHARACTERISTICS:  
        (A) LENGTH: 24 base pairs  
        (B) TYPE: nucleic acid  
        (C) STRANDEDNESS: single  
10 (D) TOPOLOGY: unknown  
    (ii) MOLECULE TYPE: DNA (genomic)  
    (iii) HYPOTHETICAL: NO  
    (iv) ANTI-SENSE: YES  
    (xi) SEQUENCE DESCRIPTION: SEQ ID NO:29:  
15 GAGTTAAATG CCTACACTGT ATCT 24  
(2) INFORMATION FOR SEQ ID NO:30:  
    (i) SEQUENCE CHARACTERISTICS:  
        (A) LENGTH: 24 base pairs  
        (B) TYPE: nucleic acid  
20 (C) STRANDEDNESS: single  
        (D) TOPOLOGY: unknown  
    (ii) MOLECULE TYPE: DNA (genomic)  
    (iii) HYPOTHETICAL: NO  
    (iv) ANTI-SENSE: NO  
25 (xi) SEQUENCE DESCRIPTION: SEQ ID NO:30:  
CAGGACTCAG AAGCTGCTAT CGAA 24  
(2) INFORMATION FOR SEQ ID NO:31:  
    (i) SEQUENCE CHARACTERISTICS:  
        (A) LENGTH: 24 base pairs  
30 (B) TYPE: nucleic acid  
        (C) STRANDEDNESS: single  
        (D) TOPOLOGY: unknown  
    (ii) MOLECULE TYPE: DNA (genomic)  
  
    (iii) HYPOTHETICAL: NO  
35 (iv) ANTI-SENSE: NO  
    (xi) SEQUENCE DESCRIPTION: SEQ ID NO:31:  
CTGCACGTGC CCTGTAGGAT TTGT 24  
(2) INFORMATION FOR SEQ ID NO:32:  
    (i) SEQUENCE CHARACTERISTICS:  
        (A) LENGTH: 23 base pairs  
40 (B) TYPE: nucleic acid  
        (C) STRANDEDNESS: single  
        (D) TOPOLOGY: unknown  
    (ii) MOLECULE TYPE: DNA (genomic)  
45 (iii) HYPOTHETICAL: NO  
    (iv) ANTI-SENSE: NO  
    (xi) SEQUENCE DESCRIPTION: SEQ ID NO:32:  
CCAGAYTGGA ARWKNAGGAA AGA 23  
(2) INFORMATION FOR SEQ ID NO:33:  
50 (i) SEQUENCE CHARACTERISTICS:  
        (A) LENGTH: 23 base pairs  
        (B) TYPE: nucleic acid  
        (C) STRANDEDNESS: single

-86-

(D) TOPOLOGY: unknown  
(ii) MOLECULE TYPE: DNA (genomic)  
(iii) HYPOTHETICAL: NO  
(iv) ANTI-SENSE: NO  
(xi) SEQUENCE DESCRIPTION: SEQ ID NO:33:  
5 CCAGAYTGGA ARWKNAGAAA AGA 23  
(2) INFORMATION FOR SEQ ID NO:34:  
(i) SEQUENCE CHARACTERISTICS:  
10 (A) LENGTH: 23 base pairs  
(B) TYPE: nucleic acid  
(C) STRANDEDNESS: single  
(D) TOPOLOGY: unknown  
(ii) MOLECULE TYPE: DNA (genomic)  
15 (iii) HYPOTHETICAL: NO  
(iv) ANTI-SENSE: NO  
(xi) SEQUENCE DESCRIPTION: SEQ ID NO:34:  
CCAGAYTGGA ARWKNAGGAA GGA 23  
(2) INFORMATION FOR SEQ ID NO:35:  
20 (i) SEQUENCE CHARACTERISTICS:  
(A) LENGTH: 23 base pairs  
(B) TYPE: nucleic acid  
(C) STRANDEDNESS: single  
(D) TOPOLOGY: unknown  
25 (ii) MOLECULE TYPE: DNA (genomic)  
(iii) HYPOTHETICAL: NO  
(iv) ANTI-SENSE: NO  
(xi) SEQUENCE DESCRIPTION: SEQ ID NO:35:  
CCAGAYTGGA ARWKNAGAAA GGA 23  
(2) INFORMATION FOR SEQ ID NO:36:  
30 (i) SEQUENCE CHARACTERISTICS:  
(A) LENGTH: 84 base pairs  
(B) TYPE: nucleic acid  
(C) STRANDEDNESS: single  
(D) TOPOLOGY: unknown  
35 (ii) MOLECULE TYPE: DNA (genomic)  
(iii) HYPOTHETICAL: NO  
(iv) ANTI-SENSE: NO  
(xi) SEQUENCE DESCRIPTION: SEQ ID NO:36:  
CAGAGTGGGC AGAGAGGCTN TTGTTAAGGG GAAATTAATC GACGTGGAAA 60  
40 AGGAAGGAAA 84  
AGTCGNTATT CCTCCAAGGG AATA  
(2) INFORMATION FOR SEQ ID NO:37:  
45 (i) SEQUENCE CHARACTERISTICS:  
(A) LENGTH: 27 amino acids  
(B) TYPE: amino acid  
(D) TOPOLOGY: unknown  
(ii) MOLECULE TYPE: peptide  
50 (iii) HYPOTHETICAL: YES  
(iv) ANTI-SENSE: NO  
(v) FRAGMENT TYPE: internal  
(xi) SEQUENCE DESCRIPTION: SEQ ID NO:37:  
Glu Trp Ala Glu Arg Leu Leu Leu Arg Gly Asn Xaa Ser Lys Trp Lys  
1 5 10 15

-87-

Arg Lys Glu Lys Ser Xaa Phe Leu Gln Gly Asn  
20 25

## (2) INFORMATION FOR SEQ ID NO:38:

- (i) SEQUENCE CHARACTERISTICS:  
(A) LENGTH: 27 amino acids  
(B) TYPE: amino acid  
(D) TOPOLOGY: unknown

- (ii) MOLECULE TYPE: peptide  
(iii) HYPOTHETICAL: YES

- (iv) ANTI-SENSE: NO

- (v) FRAGMENT TYPE: internal

- (xi) SEQUENCE DESCRIPTION: SEQ ID NO:38:

Arg Val Gly Arg Glu Ala Xaa Val Lys Gly Lys Leu Ile Glu Val Glu  
1 5 10 15  
Lys Glu Gly Lys Val Xaa Ile Pro Pro Arg Glu  
20 25

## (2) INFORMATION FOR SEQ ID NO:39:

- (i) SEQUENCE CHARACTERISTICS:  
(A) LENGTH: 28 amino acids  
(B) TYPE: amino acid  
(D) TOPOLOGY: unknown

- (ii) MOLECULE TYPE: peptide

- (iii) HYPOTHETICAL: YES

- (iv) ANTI-SENSE: NO

- (v) FRAGMENT TYPE: internal

- (xi) SEQUENCE DESCRIPTION: SEQ ID NO:39:

Gln Ser Gly Gln Arg Gly Xaa Cys Xaa Gly Glu Ile Asn Arg Ser Gly  
1 5 10 15  
Lys Gly Arg Lys Ser Arg Tyr Ser Ser Lys Gly Leu  
20 25

## (2) INFORMATION FOR SEQ ID NO:40:

- (i) SEQUENCE CHARACTERISTICS:  
(A) LENGTH: 129 base pairs  
(B) TYPE: nucleic acid  
(C) STRANDEDNESS: single  
(D) TOPOLOGY: unknown

- (ii) MOLECULE TYPE: DNA (genomic)

- (iii) HYPOTHETICAL: NO

- (iv) ANTI-SENSE: NO

- (xi) SEQUENCE DESCRIPTION: SEQ ID NO:40:

CTGCCCCACTC TGAGGTCATA ACCTGCTGGT TGGAGCCATT CTTTCAGAAAA TGGCTCTATA 60  
AGTATTTCTT TTCTGATTTT CCAGTCTGGA AGTAGCATTT TACCACCGAA ACCTTTATTT 120  
TTAATTTAA 129

## (2) INFORMATION FOR SEQ ID NO:41:

- (i) SEQUENCE CHARACTERISTICS:  
(A) LENGTH: 42 amino acids  
(B) TYPE: amino acid  
(D) TOPOLOGY: unknown

- (ii) MOLECULE TYPE: peptide

- (iii) HYPOTHETICAL: NO

- (iv) ANTI-SENSE: NO

- (v) FRAGMENT TYPE: N-terminal

- (xi) SEQUENCE DESCRIPTION: SEQ ID NO:41:

Xaa Ile Lys Asn Lys Gly Phe Gly Gly Lys Met Leu Leu Pro Asp Trp

-88-

1 5 10 15  
 Lys Ile Arg Lys Glu Ile Leu Ile Glu Pro Phe Ser Glu Glu Trp Leu  
 20 25 30  
 Gln Pro Ala Gly Tyr Asp Leu Arg Val Gly  
 35 40

## (2) INFORMATION FOR SEQ ID NO:42:

## (i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 740 base pairs  
 (B) TYPE: nucleic acid  
 (C) STRANDEDNESS: single  
 (D) TOPOLOGY: unknown

## (ii) MOLECULE TYPE: DNA (genomic)

## (iii) HYPOTHETICAL: NO

## (iv) ANTI-SENSE: NO

## (xi) SEQUENCE DESCRIPTION: SEQ ID NO:42:

TCCTCCAAGG GAATACGCCT TAATCCTAAC CCTCGAGAGG ATAAAGTTGC CCGACGATGT 60  
 TATGGGGGAT ATGAAGATAA GGAGCAGTTT AGCAAGAGAA GGGGTTATTG GTTCTTTTGC 120  
 TTGGGTTGAC CCAGGATGGG ATGGAAACTT AACACTAATG CTCTACAATG CCTCAAATGA 180  
 ACCTGTCGAA TTAAGATATG GAGAGAGATT TGTGCAGATC GCATTTATAA GGCTAGAGGG 240  
 TCCGGCAAGA AACCCCTTACA GAGGAAACTA TCAGGGGAGC ACAAGGTTAG CGTTTTCAAA 300  
 GAGAAAGAAA CTCTAGCGTC TTTTCAATAG CATCCTCAAT ATCTCGTGTG AAGTAATCAA 360  
 TGTAATAACT TGCTGGGTGG GTTTTTAGGG ATTCAAATC GTAAGATGGG CCTGTATAGC 420  
 AGAAAACTAT TTTTGCCTCT TCTTCATTTA TCTTTCTGTG AATAAAAAAT CCAACATCCA 480  
 CACTAGTTCC AAAAGATATT GTTTGCGTGA TTACCAACAA GATCTTGGCA TTATTTTTGA 540  
 TCTTATACTC TATTCTCCTT TCTCCCTCCA ATTTGCCCAA AATAAACCTG GGTAGTATAC 600  
 ATTCACCTCT CTCTTTTAAA TTCCTATAAA TTCGTACATA GTTTAGAAAA ATGTCAAATT 660  
 CTTTNTTCCC TGTAAATTA ACCNCNAAAT CTTTATNANN AANCTTTTTA TAATTCCTCAA 720  
 AACCCTAAT TTTCCCCTTN 740

## (2) INFORMATION FOR SEQ ID NO:43:

## (i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 246 amino acids  
 (B) TYPE: amino acid  
 (C) STRANDEDNESS: unknown  
 (D) TOPOLOGY: unknown

## (ii) MOLECULE TYPE: peptide

## (iii) HYPOTHETICAL: YES

## (iv) ANTI-SENSE: NO

## (v) FRAGMENT TYPE: N-terminal

## (xi) SEQUENCE DESCRIPTION: SEQ ID NO:43:

Leu Gln Gly Asn Thr Pro Xaa Ser Xaa Pro Ser Arg Gly Xaa Ser Cys  
 1 5 10 15  
 Pro Thr Met Leu Trp Gly Ile Xaa Arg Xaa Gly Ala Val Xaa Gln Glu  
 20 25 30  
 Lys Gly Leu Leu Val Leu Leu Leu Gly Leu Thr Gln Asp Gly Met Glu  
 35 40 45  
 Thr Xaa His Xaa Cys Ser Thr Met Pro Gln Met Asn Leu Ser Asn Xaa  
 50 55 60  
 Asp Met Glu Arg Asp Leu Cys Arg Ser His Leu Xaa Gly Xaa Arg Val  
 65 70 75 80  
 Arg Gln Glu Thr Leu Thr Glu Glu Thr Ile Arg Gly Ala Gln Gly Xaa  
 85 90 95  
 Arg Phe Gln Arg Glu Arg Asn Ser Ser Val Phe Ser Ile Ala Ser Ser  
 100 105 110  
 Ile Ser Arg Val Lys Xaa Ser Met Xaa Ile Leu Ala Gly Trp Val Phe

-89-

[illegible]

(2) INFORMATION FOR SEQ ID NO:44:

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 246 amino acids

(B) TYPE: amino acid

(D) TOPOLOGY: unknown

(ii) MOLECULE TYPE: peptide

(iii) HYPOTHETICAL: YES

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:44:

17	SEQUENCE DESCRIPTION: SEQ ID NO: 1														
Pro	Pro	Arg	Glu	Tyr	Ala	Leu	Ile	Leu	Thr	Leu	Glu	Arg	Ile	Lys	Leu
1				5					10					15	
Pro	Asn	Asn	Val	Met	Gly	Asp	Met	Lys	Ile	Arg	Ser	Ser	Leu	Ala	Arg
			20					25					30		
Glu	Gly	Val	Ile	Gly	Ser	Phe	Ala	Trp	Val	Asp	Pro	Gly	Trp	Asp	Gly
		35					40					45			
Asn	Leu	Thr	Leu	Met	Leu	Tyr	Asn	Ala	Ser	Asn	Glu	Pro	Val	Glu	Leu
	50					55					60				
Arg	Tyr	Gly	Glu	Arg	Phe	Val	Gln	Ile	Ala	Phe	Ile	Arg	Leu	Glu	Gly
65					70					75					80
Pro	Ala	Arg	Asn	Pro	Tyr	Arg	Gly	Asn	Tyr	Gln	Gly	Ser	Thr	Arg	Leu
				85					90					95	
Ala	Phe	Ser	Lys	Arg	Lys	Lys	Leu	Xaa	Arg	Leu	Phe	Asn	Ser	Ile	Leu
			100					105						110	
Asn	Ile	Ser	Cys	Glu	Val	Ile	Asn	Val	Asn	Thr	Cys	Trp	Val	Gly	Phe
		115					120					125			
Xaa	Gly	Phe	Lys	Leu	Val	Arg	Trp	Ala	Cys	Ile	Ala	Glu	Asn	Tyr	Phe
	130					135					140				
Cys	Leu	Phe	Phe	Ile	Tyr	Leu	Ser	Val	Asn	Lys	Lys	Ser	Asn	Ile	His
145					150					155					160
Thr	Ser	Ser	Lys	Arg	Tyr	Cys	Leu	Arg	Asp	Tyr	Gln	Gln	Asp	Leu	Gly
				165					170					175	
Ile	Ile	Phe	Asp	Leu	Ile	Leu	Tyr	Ser	Pro	Phe	Ser	Leu	Gln	Phe	Ala
			180				185						190		
Gln	Asn	Lys	Pro	Gly	Xaa	Tyr	Thr	Phe	Thr	Pro	Leu	Phe	Xaa	Ile	Pro
		195					200					205			
Ile	Asn	Ser	Tyr	Ile	Val	Xaa	Lys	Asn	Val	Lys	Phe	Phe	Xaa	Pro	Cys
	210					215					220				

-90-

Xaa Ile Asn Xaa Xaa Ile Phe Xaa Xaa Xaa Leu Phe Ile Ile Pro Lys  
 225 230 235 240  
 Thr Pro Asn Phe Pro Leu  
 245

5

## (2) INFORMATION FOR SEQ ID NO:45:

## (i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 246 amino acids  
 (B) TYPE: amino acid  
 (D) TOPOLOGY: unknown

10

## (ii) MOLECULE TYPE: peptide

## (iii) HYPOTHETICAL: YES

## (xi) SEQUENCE DESCRIPTION: SEQ ID NO:45:

15

Ser Ser Lys Gly Ile Arg Leu Asn Pro Asn Pro Arg Glu Asp Lys Val  
 1 5 10 15

Ala Arg Arg Cys Tyr Gly Gly Tyr Glu Asp Lys Glu Gln Phe Ser Lys  
 20 25 30

Arg Arg Gly Tyr Trp Phe Phe Cys Leu Gly Xaa Pro Arg Met Gly Trp  
 35 40 45

20

Lys Leu Asn Thr Asn Ala Leu Gln Cys Leu Lys Xaa Thr Cys Arg Ile  
 50 55 60

Lys Ile Trp Arg Glu Ile Cys Ala Asp Arg Ile Tyr Lys Ala Arg Gly  
 65 70 75 80

Ser Gly Lys Lys Pro Leu Gln Arg Lys Leu Ser Gly Glu His Lys Val  
 85 90 95

25

Ser Val Phe Lys Glu Lys Glu Thr Leu Ala Ser Phe Gln Xaa His Pro  
 100 105 110

Gln Tyr Leu Val Xaa Ser Asn Gln Cys Lys Tyr Leu Leu Gly Gly Phe  
 115 120 125

30

Leu Gly Ile Gln Thr Arg Lys Met Gly Leu Tyr Ser Arg Lys Leu Phe  
 130 135 140

Leu Pro Leu Leu His Leu Ser Phe Cys Glu Xaa Lys Ile Gln His Pro  
 145 150 155 160

His Xaa Phe Gln Lys Ile Leu Phe Ala Xaa Leu Pro Thr Arg Ser Trp  
 165 170 175

35

His Tyr Phe Xaa Ser Tyr Thr Leu Phe Ser Phe Leu Pro Pro Ile Cys  
 180 185 190

Pro Lys Xaa Thr Trp Val Val Tyr Ile His Ser Ser Leu Leu Asn Ser  
 195 200 205

40

Tyr Lys Phe Val His Ser Leu Glu Lys Cys Gln Ile Leu Xaa Ser Leu  
 210 215 220

Leu Asn Xaa Pro Xaa Asn Leu Tyr Xaa Xaa Xaa Phe Tyr Asn Ser Gln  
 225 230 235 240

Asn Pro Xaa Phe Ser Pro  
 245

45

## (2) INFORMATION FOR SEQ ID NO:46:

## (i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 31 amino acids  
 (B) TYPE: amino acid  
 (C) STRANDEDNESS: unknown  
 (D) TOPOLOGY: unknown

50

## (ii) MOLECULE TYPE: peptide

## (iii) HYPOTHETICAL: NO

## (iv) ANTI-SENSE: NO

## (v) FRAGMENT TYPE: N-terminal

-91-

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:46:  
 Met Leu His His Val Lys Leu Ile Tyr Ala Thr Lys Ser Arg Lys Leu  
 1 5 10 15  
 Val Gly Lys Lys Ile Val Xaa Xaa Xaa Pro Gly Ser Ile Ala Ala  
 20 25 30

## (2) INFORMATION FOR SEQ ID NO:47:

## (i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 17 amino acids  
 (B) TYPE: amino acid  
 (C) STRANDEDNESS: unknown  
 (D) TOPOLOGY: unknown

(ii) MOLECULE TYPE: peptide

(iii) HYPOTHETICAL: NO

(iv) ANTI-SENSE: NO

(v) FRAGMENT TYPE: internal

## (xi) SEQUENCE DESCRIPTION: SEQ ID NO:47:

Lys Tyr Asp Val Val Ile Met Ala Ala Ala Val Ser Asp Phe Arg Phe  
 1 5 10 15  
 Lys

## (2) INFORMATION FOR SEQ ID NO:48:

## (i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 24 amino acids  
 (B) TYPE: amino acid  
 (C) STRANDEDNESS: unknown  
 (D) TOPOLOGY: unknown

(ii) MOLECULE TYPE: peptide

(iii) HYPOTHETICAL: NO

(iv) ANTI-SENSE: NO

(v) FRAGMENT TYPE: internal

## (xi) SEQUENCE DESCRIPTION: SEQ ID NO:48:

Ala Asp Leu Val Val Gly Asn Thr Leu Glu Ala Phe Gly Ser Glu Glu  
 1 5 10 15  
 Asn Gln Val Val Leu Ile Gly Arg  
 20

## (2) INFORMATION FOR SEQ ID NO:49:

## (i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 21 base pairs  
 (B) TYPE: nucleic acid  
 (C) STRANDEDNESS: single  
 (D) TOPOLOGY: unknown

(ii) MOLECULE TYPE: DNA (genomic)

(iii) HYPOTHETICAL: NO

(iv) ANTI-SENSE: NO

## (xi) SEQUENCE DESCRIPTION: SEQ ID NO:49:

CTATTGAGTA CGAACGCCAT C

21

## (2) INFORMATION FOR SEQ ID NO:50:

## (i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 20 base pairs  
 (B) TYPE: nucleic acid  
 (C) STRANDEDNESS: single  
 (D) TOPOLOGY: unknown

(ii) MOLECULE TYPE: DNA (genomic)

(iii) HYPOTHETICAL: NO

(iv) ANTI-SENSE: NO

-92-

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:50:

GTCACGCTTG CTCCACTCCG

20

(2) INFORMATION FOR SEQ ID NO:51:

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 437 amino acids

(B) TYPE: amino acid

(D) TOPOLOGY: unknown

(ii) MOLECULE TYPE: protein

(iii) HYPOTHETICAL: NO

(iv) ANTI-SENSE: NO

(vi) ORIGINAL SOURCE:

(A) ORGANISM: Methanococcus Jannaschii

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:51:

Met	Ile	Ser	Glu	Ile	Met	His	Pro	Thr	Lys	Leu	Leu	Lys	Gly	Thr	Lys
1				5					10					15	
Ser	Lys	Leu	Leu	Glu	Asn	Lys	Lys	Ile	Leu	Val	Ala	Val	Thr	Ser	Ser
			20					25					30		
Ile	Ala	Ala	Ile	Glu	Thr	Pro	Lys	Leu	Met	Arg	Glu	Leu	Ile	Arg	His
		35					40					45			
Gly	Ala	Glu	Val	Tyr	Cys	Ile	Ile	Thr	Glu	Glu	Thr	Lys	Lys	Ile	Ile
	50					55					60				
Gly	Lys	Glu	Ala	Leu	Lys	Phe	Gly	Cys	Gly	Asn	Glu	Val	Tyr	Glu	Glu
65				70					75					80	
Ile	Thr	Gly	Xaa	Xaa	Xaa	Xaa	Asp	Ile	Glu	His	Ile	Leu	Leu	Tyr	
			85					90					95		
Xaa	Xaa	Xaa	Xaa	Asn	Glu	Cys	Asp	Cys	Leu	Leu	Ile	Tyr	Pro	Ala	Thr
			100					105					110		
Ala	Asn	Ile	Ile	Ser	Lys	Ile	Asn	Leu	Gly	Ile	Ala	Asp	Asn	Ile	Val
		115					120					125			
Asn	Thr	Thr	Ala	Leu	Met	Phe	Phe	Gly	Asn	Lys	Pro	Ile	Phe	Ile	Val
	130					135					140				
Pro	Ala	Met	His	Glu	Asn	Met	Phe	Asn	Xaa	Xaa	Ala	Ile	Lys	Arg	His
				150					155					160	
Ile	Asp	Lys	Leu	Lys	Glu	Lys	Asp	Lys	Ile	Tyr	Ile	Ile	Ser	Pro	Lys
			165						170					175	
Phe	Glu	Glu	Xaa	Xaa	Xaa	Xaa	Xaa	Xaa	Gly	Lys	Ala	Lys	Val	Ala	Asn
			180				185						190		
Ile	Glu	Asp	Val	Val	Lys	Ala	Val	Ile	Glu	Lys	Ile	Gly	Asn	Asn	Leu
	195						200					205			
Lys	Lys	Glu	Gly	Asn	Arg	Val	Leu	Ile	Leu	Asn	Gly	Gly	Thr	Val	Glu
	210					215					220				
Phe	Ile	Asp	Lys	Val	Arg	Val	Ile	Ser	Asn	Leu	Ser	Ser	Gly	Lys	Met
225				230						235				240	
Gly	Val	Ala	Leu	Ala	Glu	Ala	Phe	Cys	Lys	Glu	Gly	Phe	Tyr	Val	Glu
			245						250					255	
Val	Ile	Thr	Ala	Met	Gly	Leu	Glu	Pro	Pro	Tyr	Tyr	Ile	Lys	Asn	His
			260					265					270		
Lys	Val	Leu	Thr	Ala	Lys	Glu	Met	Leu	Asn	Lys	Ala	Ile	Glu	Xaa	Xaa
		275					280					285			
Leu	Xaa	Ala	Lys	Asp	Phe	Asp	Ile	Ile	Ile	Ser	Ser	Ala	Ala	Ile	Ser
	290					295					300				
Asp	Phe	Thr	Val	Glu	Ser	Xaa	Phe	Glu	Gly	Lys	Leu	Ser	Ser	Glu	Glu
305				310						315				320	
Glu	Xaa	Xaa	Xaa	Xaa	Leu	Ile	Leu	Lys	Leu	Lys	Arg	Xaa	Asn	Pro	Lys



-93-

325 330 335  
 Val Leu Glu Glu Leu Arg Arg Ile Tyr Lys Asp Xaa Lys Val Ile Ile  
 340 345 350  
 Gly Phe Lys Ala Glu Tyr Asn Leu Asp Glu Lys Glu Leu Ile Asn Arg  
 355 360 365  
 Ala Lys Glu Arg Leu Asn Lys Tyr Asn Leu Asn Met Ile Ile Ala Asn  
 370 375 380  
 Asp Leu Ser Lys Xaa Xaa His Tyr Phe Gly Asp Asp Tyr Ile Glu Val  
 385 390 395 400  
 Tyr Ile Ile Thr Lys Tyr Glu Val Glu Lys Ile Ser Gly Ser Lys Lys  
 405 410 415  
 Xaa Glu Ile Ser Glu Arg Ile Val Glu Lys Val Lys Lys Leu Val Lys  
 420 425 430  
 Ser Xaa Xaa Xaa Xaa  
 435

## (2) INFORMATION FOR SEQ ID NO:52:

## (i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 444 amino acids  
 (B) TYPE: amino acid  
 (D) TOPOLOGY: unknown

## (ii) MOLECULE TYPE: protein

## (iii) HYPOTHETICAL: NO

## (iv) ANTI-SENSE: NO

## (vi) ORIGINAL SOURCE:

- (A) ORGANISM: Escherichia coli

## (xi) SEQUENCE DESCRIPTION: SEQ ID NO:52:

Met Lys Ala Arg Gln Gln Lys Tyr Cys Asp Lys Ile Ala Asn Phe Trp  
 1 5 10 15  
 Cys His Pro Thr Gly Lys Ile Ile Met Ser Leu Ala Gly Lys Lys Ile  
 20 25 30  
 Val Leu Gly Val Ser Gly Gly Ile Ala Ala Tyr Lys Thr Pro Glu Leu  
 35 40 45  
 Val Arg Arg Leu Arg Asp Arg Gly Ala Asp Val Arg Val Ala Met Thr  
 50 55 60  
 Glu Ala Ala Lys Ala Phe Ile Thr Pro Leu Ser Leu Gln Ala Val Ser  
 65 70 75 80  
 Gly Tyr Pro Val Ser Asp Ser Leu Leu Asp Pro Ala Ala Glu Ala Ala  
 85 90 95  
 Met Gly His Ile Glu Leu Gly Xaa Xaa Xaa Xaa Lys Trp Ala Asp Leu  
 100 105 110  
 Val Ile Leu Ala Pro Ala Thr Ala Asp Leu Ile Ala Arg Val Ala Ala  
 115 120 125  
 Gly Met Ala Asn Asp Leu Val Ser Thr Ile Cys Leu Ala Thr Pro Xaa  
 130 135 140  
 Xaa Ala Pro Val Ala Val Leu Pro Ala Met Asn Gln Gln Met Tyr Arg  
 145 150 155 160  
 Ala Ala Ala Thr Gln His Asn Leu Glu Val Leu Ala Xaa Ser Arg Gly  
 165 170 175  
 Leu Leu Ile Trp Gly Pro Asp Ser Gly Ser Gln Ala Cys Gly Asp Ile  
 180 185 190  
 Gly Pro Gly Arg Xaa Xaa Asp Pro Leu Thr Ile Val Asp Met Ala Val  
 195 200 205  
 Ala His Phe Ser Pro Val Asn Asp Leu Lys His Leu Asn Ile Met Ile  
 210 215 220

-94-

Thr Ala Gly Pro Thr Arg Glu Pro Leu Asp Pro Val Arg Tyr Ile Ser  
 225 230 235 240  
 Asn His Ser Ser Gly Lys Met Gly Phe Ala Ile Ala Ala Ala Ala  
 245 250 255  
 5 Arg Arg Gly Ala Asn Val Thr Leu Val Ser Gly Pro Val Ser Leu Pro  
 260 265 270  
 Thr Pro Pro Phe Val Lys Arg Val Asp Val Met Thr Ala Leu Glu Met  
 275 280 285  
 10 Glu Ala Ala Val Asn Xaa Xaa Ala Ser Val Gln Gln Gln Asn Ile Phe  
 290 295 300  
 Ile Gly Cys Ala Ala Val Ala Asp Tyr Arg Ala Ala Thr Val Ala Pro  
 305 310 315 320  
 Glu Lys Ile Lys Lys Gln Ala Thr Gln Gly Asp Glu Leu Thr Ile Lys  
 325 330 335  
 15 Met Val Lys Xaa Asn Pro Asp Ile Val Ala Gly Val Ala Ala Leu Lys  
 340 345 350  
 Asp His Arg Pro Tyr Val Val Gly Phe Ala Ala Glu Thr Asn Asn Xaa  
 355 360 365  
 20 Xaa Xaa Xaa Val Glu Glu Tyr Ala Arg Gln Lys Arg Ile Arg Lys Asn  
 370 375 380  
 Leu Asp Leu Ile Cys Ala Asn Asp Val Ser Gln Pro Thr Gln Gly Phe  
 385 390 395 400  
 Asn Ser Asp Asn Asn Ala Leu His Leu Phe Trp Gln Asp Gly Asp Lys  
 405 410 415  
 25 Val Leu Pro Leu Glu Arg Lys Glu Leu Gly Gln Leu Leu Leu Asp  
 420 425 430  
 Glu Ile Val Thr Arg Tyr Asp Glu Lys Asn Arg Arg  
 435 440

## (2) INFORMATION FOR SEQ ID NO:53:

## (i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 14 amino acids  
 (B) TYPE: amino acid  
 (D) TOPOLOGY: unknown

## (ii) MOLECULE TYPE: peptide

## (iii) HYPOTHETICAL: YES

## (iv) ANTI-SENSE: NO

## (v) FRAGMENT TYPE: internal

## (xi) SEQUENCE DESCRIPTION: SEQ ID NO:53:

Xaa Gly Xaa Xaa Asp Xaa Xaa Xaa Xaa Gly Xaa Xaa Xaa Xaa  
 1 5 10

## (2) INFORMATION FOR SEQ ID NO:54:

## (i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 14 amino acids  
 (B) TYPE: amino acid  
 (D) TOPOLOGY: unknown

## (ii) MOLECULE TYPE: peptide

## (iii) HYPOTHETICAL: NO

## (iv) ANTI-SENSE: NO

## (v) FRAGMENT TYPE: internal

## (xi) SEQUENCE DESCRIPTION: SEQ ID NO:54:

Phe Ala Trp Val Asp Pro Gly Trp Asp Gly Asn Thr Leu Met  
 1 5 10

-95-

## (2) INFORMATION FOR SEQ ID NO:55:

## (i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 14 amino acids

(B) TYPE: amino acid

(D) TOPOLOGY: unknown

(ii) MOLECULE TYPE: peptide

(iii) HYPOTHETICAL: NO

(iv) ANTI-SENSE: NO

(v) FRAGMENT TYPE: internal

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:55:

Ala Gly Trp Ile Asp Ala Gly Phe Lys Gly Lys Ile Thr Leu

1

5

10

## (2) INFORMATION FOR SEQ ID NO:56:

## (i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 14 amino acids

(B) TYPE: amino acid

(D) TOPOLOGY: unknown

(ii) MOLECULE TYPE: peptide

(iii) HYPOTHETICAL: NO

(iv) ANTI-SENSE: NO

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:56:

Ser Ala Val His Asp Pro Gly Tyr Glu Gly Arg Pro Glu Tyr

1

5

10

## (2) INFORMATION FOR SEQ ID NO:57:

## (i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 14 amino acids

(B) TYPE: amino acid

(D) TOPOLOGY: unknown

(ii) MOLECULE TYPE: peptide

(iii) HYPOTHETICAL: NO

(iv) ANTI-SENSE: NO

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:57:

Pro Thr Ile Val Asp Ala Gly Phe Glu Gly Gln Leu Thr Ile

1

5

10

## (2) INFORMATION FOR SEQ ID NO:58:

## (i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 14 amino acids

(B) TYPE: amino acid

(D) TOPOLOGY: unknown

(ii) MOLECULE TYPE: peptide

(iii) HYPOTHETICAL: NO

(iv) ANTI-SENSE: NO

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:58:

Ala His Arg Ile Asp Pro Gly Trp Ser Gly Cys Ile Val Leu

1

5

10

## (2) INFORMATION FOR SEQ ID NO:59:

## (i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 24 base pairs

(B) TYPE: nucleic acid

(C) STRANDEDNESS: single

(D) TOPOLOGY: unknown

(ii) MOLECULE TYPE: DNA (genomic)

(iii) HYPOTHETICAL: NO

(iv) ANTI-SENSE: YES

-96-

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:59:  
GAGTTAAATG CCTACACTGT ATCT 24  
(2) INFORMATION FOR SEQ ID NO:60:  
(i) SEQUENCE CHARACTERISTICS:  
5 (A) LENGTH: 24 base pairs  
(B) TYPE: nucleic acid  
(C) STRANDEDNESS: single  
(D) TOPOLOGY: unknown  
(ii) MOLECULE TYPE: DNA (genomic)  
10 (iii) HYPOTHETICAL: NO  
(iv) ANTI-SENSE: NO  
(xi) SEQUENCE DESCRIPTION: SEQ ID NO:60:  
CAGGACTCAG AAGCTGCTAT CGAA 24  
(2) INFORMATION FOR SEQ ID NO:61:  
15 (i) SEQUENCE CHARACTERISTICS:  
(A) LENGTH: 24 base pairs  
(B) TYPE: nucleic acid  
(C) STRANDEDNESS: single  
(D) TOPOLOGY: unknown  
20 (ii) MOLECULE TYPE: DNA (genomic)  
(iii) HYPOTHETICAL: NO  
(iv) ANTI-SENSE: NO  
(xi) SEQUENCE DESCRIPTION: SEQ ID NO:61:  
25 CTGCACGTGC CCTGTAGGAT TTGT 24  
(2) INFORMATION FOR SEQ ID NO:62:  
(i) SEQUENCE CHARACTERISTICS:  
(A) LENGTH: 21 base pairs  
(B) TYPE: nucleic acid  
(C) STRANDEDNESS: single  
30 (D) TOPOLOGY: linear  
(ii) MOLECULE TYPE: DNA (genomic)  
(iii) HYPOTHETICAL: NO  
(iv) ANTI-SENSE: NO  
(xi) SEQUENCE DESCRIPTION: SEQ ID NO:62:  
35 CTATTGAGTA CGAACGCCAT C 21  
(2) INFORMATION FOR SEQ ID NO:63:  
(i) SEQUENCE CHARACTERISTICS:  
(A) LENGTH: 20 base pairs  
(B) TYPE: nucleic acid  
40 (C) STRANDEDNESS: single  
(D) TOPOLOGY: linear  
(ii) MOLECULE TYPE: DNA (genomic)  
(iii) HYPOTHETICAL: NO  
(iv) ANTI-SENSE: YES  
45 (xi) SEQUENCE DESCRIPTION: SEQ ID NO:63:  
GTCACGCTTG CTCCACTCCG 20  
(2) INFORMATION FOR SEQ ID NO:64:  
(i) SEQUENCE CHARACTERISTICS:  
50 (A) LENGTH: 23 base pairs  
(B) TYPE: nucleic acid  
(C) STRANDEDNESS: single  
(D) TOPOLOGY: unknown  
(ii) MOLECULE TYPE: DNA (genomic)  
(iii) HYPOTHETICAL: NO

-97-

(iv) ANTI-SENSE: NO

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:64:

GAGGAGAGCA GGAAAGGTGG AAC

23

(2) INFORMATION FOR SEQ ID NO:65:

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 21 base pairs

(B) TYPE: nucleic acid

(C) STRANDEDNESS: single

(D) TOPOLOGY: unknown

(ii) MOLECULE TYPE: DNA (genomic)

(iii) HYPOTHETICAL: NO

(iv) ANTI-SENSE: YES

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:65:

GCTGGGAGAA GACTTCACTG G

21

(2) INFORMATION FOR SEQ ID NO:66:

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 8 amino acids

(B) TYPE: amino acid

(C) STRANDEDNESS: unknown

(D) TOPOLOGY: unknown

(ii) MOLECULE TYPE: peptide

(iii) HYPOTHETICAL: NO

(iv) ANTI-SENSE: NO

(v) FRAGMENT TYPE: internal

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:66:

His His Val Lys Leu Ile Tyr Ala

1

5

(2) INFORMATION FOR SEQ ID NO:67:

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 8 amino acids

(B) TYPE: amino acid

(C) STRANDEDNESS: unknown

(D) TOPOLOGY: unknown

(ii) MOLECULE TYPE: peptide

(iii) HYPOTHETICAL: NO

(iv) ANTI-SENSE: NO

(v) FRAGMENT TYPE: internal

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:67:

Lys Tyr Asp Ala Val Ile Met Ala

1

5

(2) INFORMATION FOR SEQ ID NO:68:

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 7 amino acids

(B) TYPE: amino acid

(C) STRANDEDNESS: unknown

(D) TOPOLOGY: unknown

(ii) MOLECULE TYPE: peptide

(iii) HYPOTHETICAL: NO

(iv) ANTI-SENSE: NO

(v) FRAGMENT TYPE: internal

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:68:

Glu Glu Asn Gln Val Val Leu

1

5

(2) INFORMATION FOR SEQ ID NO:69:

-98-

(i) SEQUENCE CHARACTERISTICS:  
 (A) LENGTH: 8 amino acids  
 (B) TYPE: amino acid  
 (C) STRANDEDNESS: unknown  
 (D) TOPOLOGY: unknown

(ii) MOLECULE TYPE: peptide  
 (iii) HYPOTHETICAL: NO  
 (iv) ANTI-SENSE: NO  
 (v) FRAGMENT TYPE: internal

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:69:  
 Pro Asp Trp Lys Ile Arg Lys Glu  
 1 5

(2) INFORMATION FOR SEQ ID NO:70:  
 (i) SEQUENCE CHARACTERISTICS:  
 (A) LENGTH: 471 base pairs  
 (B) TYPE: nucleic acid  
 (C) STRANDEDNESS: double  
 (D) TOPOLOGY: unknown

(ii) MOLECULE TYPE: DNA (genomic)  
 (iii) HYPOTHETICAL: NO  
 (iv) ANTI-SENSE: NO  
 (xi) SEQUENCE DESCRIPTION: SEQ ID NO:70:  
 ATGCTACTTC CAGACTGGAA AATCAGAAAA GAAATACTTA TAGAGCCATT TTCTGAAGAA 60  
 TCGCTCCAAC CAGCAGGTTA TGACCTCAGA GTGGGCAGAG AGGCTTTTGT TAAGGGGAAA 120  
 TTAATCGACG TGGAAAAGGA AGGAAAAGTC GTTATTCCTC CAAGGGAATA CGCCTTAATC 180  
 CTAACCCCTCG AGAGGATAAA GTTGCCCGAC GATGTTATGG GGGATATGAA GATAAGGAGC 240  
 AGTTTAGCAA GAGAAGGGGT TATTGGTTCT TTTGCTTGGG TTGACCCAGG ATGGGATGGA 300  
 AACTTAACAC TAATGCTCTA CAATGCCTCA AATGAACCTG TCGAATTAAG ATATGGAGAG 360  
 AGATTTGTGC AGATCGCATT TATAAGGCTA GAGGGTCCGG CAAGAAACCC TTACAGAGGA 420  
 AACTATCAGG GGAGCACAAG GTTAGCGTTT TCAAAGAGAA AGAAACTCTA G 471

(2) INFORMATION FOR SEQ ID NO:71:  
 (i) SEQUENCE CHARACTERISTICS:  
 (A) LENGTH: 156 amino acids  
 (B) TYPE: amino acid  
 (C) STRANDEDNESS: unknown  
 (D) TOPOLOGY: unknown

(ii) MOLECULE TYPE: protein  
 (iii) HYPOTHETICAL: NO  
 (iv) ANTI-SENSE: NO  
 (xi) SEQUENCE DESCRIPTION: SEQ ID NO:71:  
 Met Leu Leu Pro Asp Trp Lys Ile Arg Lys Glu Ile Leu Ile Glu Pro  
 1 5 10 15  
 Phe Ser Glu Glu Ser Leu Gln Pro Ala Gly Tyr Asp Leu Arg Val Gly  
 20 25 30  
 Arg Glu Ala Phe Val Lys Gly Lys Leu Ile Asp Val Glu Lys Glu Gly  
 35 40 45  
 Lys Val Val Ile Pro Pro Arg Glu Tyr Ala Leu Ile Leu Thr Leu Glu  
 50 55 60  
 Arg Ile Lys Leu Pro Asp Asp Val Met Gly Asp Met Lys Ile Arg Ser  
 65 70 75 80  
 Ser Leu Ala Arg Glu Gly Val Ile Gly Ser Phe Ala Trp Val Asp Pro  
 85 90 95  
 Gly Trp Asp Gly Asn Leu Thr Leu Met Leu Tyr Asn Ala Ser Asn Glu  
 100 105 110

-99-

Pro Val Glu Leu Arg Tyr Gly Glu Arg Phe Val Gln Ile Ala Phe Ile  
 115 120 125  
 Arg Leu Glu Gly Pro Ala Arg Asn Pro Tyr Arg Gly Asn Tyr Gln Gly  
 130 135 140  
 Ser Thr Arg Leu Ala Phe Ser Lys Arg Lys Lys Leu  
 145 150 155

## (2) INFORMATION FOR SEQ ID NO:72:

- (i) SEQUENCE CHARACTERISTICS:  
 (A) LENGTH: 13 amino acids  
 (B) TYPE: amino acid  
 (C) STRANDEDNESS: unknown  
 (D) TOPOLOGY: unknown.

(ii) MOLECULE TYPE: peptide

(iii) HYPOTHETICAL: YES

(iv) ANTI-SENSE: NO

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:72:

Xaa Gly Xaa Xaa Asp Xaa Xaa Xaa Gly Xaa Xaa Xaa Xaa  
 1 5 10

## (2) INFORMATION FOR SEQ ID NO:73:

- (i) SEQUENCE CHARACTERISTICS:  
 (A) LENGTH: 14 amino acids  
 (B) TYPE: amino acid  
 (C) STRANDEDNESS: unknown  
 (D) TOPOLOGY: unknown

(ii) MOLECULE TYPE: peptide

(iii) HYPOTHETICAL: NO

(iv) ANTI-SENSE: NO

(v) FRAGMENT TYPE: internal

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:73:

Phe Ala Trp Val Asp Pro Gly Trp Asp Gly Asn Thr Leu Met  
 1 5 10

## (2) INFORMATION FOR SEQ ID NO:74:

- (i) SEQUENCE CHARACTERISTICS:  
 (A) LENGTH: 14 amino acids  
 (B) TYPE: amino acid  
 (C) STRANDEDNESS: unknown  
 (D) TOPOLOGY: unknown

(ii) MOLECULE TYPE: peptide

(iii) HYPOTHETICAL: NO

(iv) ANTI-SENSE: NO

(v) FRAGMENT TYPE: internal

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:74:

Ala Gly Trp Ile Asp Ala Gly Phe Lys Gly Lys Ile Thr Leu  
 1 5 10

## (2) INFORMATION FOR SEQ ID NO:75:

- (i) SEQUENCE CHARACTERISTICS:  
 (A) LENGTH: 14 amino acids  
 (B) TYPE: amino acid  
 (C) STRANDEDNESS: unknown  
 (D) TOPOLOGY: unknown

(ii) MOLECULE TYPE: peptide

(iii) HYPOTHETICAL: NO

(iv) ANTI-SENSE: NO

-100-

(v) FRAGMENT TYPE: internal

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:75:

Ser Ala Val His Asp Pro Gly Tyr Glu Gly Arg Pro Glu Tyr

1 5 10

(2) INFORMATION FOR SEQ ID NO:76:

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 14 amino acids

(B) TYPE: amino acid

(C) STRANDEDNESS: unknown

(D) TOPOLOGY: unknown

(ii) MOLECULE TYPE: peptide

(iii) HYPOTHETICAL: NO

(iv) ANTI-SENSE: NO

(v) FRAGMENT TYPE: internal

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:76:

Pro Thr Ile Val Asp Ala Gly Phe Glu Gly Gln Leu Thr Ile

1 5 10

(2) INFORMATION FOR SEQ ID NO:77:

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 14 amino acids

(B) TYPE: amino acid

(C) STRANDEDNESS: unknown

(D) TOPOLOGY: unknown

(ii) MOLECULE TYPE: peptide

(iii) HYPOTHETICAL: NO

(iv) ANTI-SENSE: NO

(v) FRAGMENT TYPE: internal

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:77:

Ala His Arg Ile Asp Pro Gly Trp Ser Gly Cys Ile Val Leu

1 5 10

(2) INFORMATION FOR SEQ ID NO:78:

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 14 amino acids

(B) TYPE: amino acid

(C) STRANDEDNESS: unknown

(D) TOPOLOGY: unknown

(ii) MOLECULE TYPE: peptide

(iii) HYPOTHETICAL: NO

(iv) ANTI-SENSE: NO

(v) FRAGMENT TYPE: internal

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:78:

Val Gly Leu Ile Asp Ser Asp Tyr Gln Gly Gln Leu Met Ile

1 5 10

(2) INFORMATION FOR SEQ ID NO:79:

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 14 amino acids

(B) TYPE: amino acid

(C) STRANDEDNESS: unknown

(D) TOPOLOGY: unknown

(ii) MOLECULE TYPE: peptide

(iii) HYPOTHETICAL: NO

(iv) ANTI-SENSE: NO



-101-

(v) FRAGMENT TYPE: internal  
 (xi) SEQUENCE DESCRIPTION: SEQ ID NO:79:  
 Ala Gly Val Val Asp Arg Asp Tyr Thr Gly Glu Val Lys Val  
 1 5 10

(2) INFORMATION FOR SEQ ID NO:80:

(i) SEQUENCE CHARACTERISTICS:  
 (A) LENGTH: 14 amino acids  
 (B) TYPE: amino acid  
 (C) STRANDEDNESS: unknown  
 (D) TOPOLOGY: unknown

(ii) MOLECULE TYPE: peptide

(iii) HYPOTHETICAL: NO

(iv) ANTI-SENSE: NO

(v) FRAGMENT TYPE: internal

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:80:  
 Ala Gly Val Ile Asp Glu Asp Tyr Arg Gly Asn Val Gly Val  
 1 5 10

(2) INFORMATION FOR SEQ ID NO:81:

(i) SEQUENCE CHARACTERISTICS:  
 (A) LENGTH: 14 amino acids  
 (B) TYPE: amino acid  
 (C) STRANDEDNESS: unknown  
 (D) TOPOLOGY: unknown

(ii) MOLECULE TYPE: peptide

(iii) HYPOTHETICAL: NO

(iv) ANTI-SENSE: NO

(v) FRAGMENT TYPE: internal

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:81:  
 Thr Gly Leu Ile Asp Pro Gly Phe Gln Gly Glu Leu Lys Leu  
 1 5 10

(2) INFORMATION FOR SEQ ID NO:82:

(i) SEQUENCE CHARACTERISTICS:  
 (A) LENGTH: 33 base pairs  
 (B) TYPE: nucleic acid  
 (C) STRANDEDNESS: single  
 (D) TOPOLOGY: unknown

(ii) MOLECULE TYPE: DNA (genomic)

(iii) HYPOTHETICAL: NO

(iv) ANTI-SENSE: NO

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:82:

GACGACGACA AGATGCTACT TCCAGACTGG AAA

33

(2) INFORMATION FOR SEQ ID NO:83:

(i) SEQUENCE CHARACTERISTICS:  
 (A) LENGTH: 35 base pairs  
 (B) TYPE: nucleic acid  
 (C) STRANDEDNESS: single  
 (D) TOPOLOGY: unknown

(ii) MOLECULE TYPE: DNA (genomic)

(iii) HYPOTHETICAL: NO

(iv) ANTI-SENSE: NO

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:83:

GGAACAAGAC CCGTCCCACT TTCACAGATG AAGAG

35

(2) INFORMATION FOR SEQ ID NO:84:

(i) SEQUENCE CHARACTERISTICS:

-102-

(A) LENGTH: 23 base pairs  
(B) TYPE: nucleic acid  
(C) STRANDEDNESS: single  
(D) TOPOLOGY: unknown

5 (ii) MOLECULE TYPE: DNA (genomic)  
(iii) HYPOTHETICAL: NO  
(iv) ANTI-SENSE: NO  
(xi) SEQUENCE DESCRIPTION: SEQ ID NO:84:  
GAGGAGAGCA GGAAAGGTGG AAC 23

10 (2) INFORMATION FOR SEQ ID NO:85:  
(i) SEQUENCE CHARACTERISTICS:  
(A) LENGTH: 23 base pairs  
(B) TYPE: nucleic acid  
(C) STRANDEDNESS: single  
15 (D) TOPOLOGY: unknown  
(ii) MOLECULE TYPE: DNA (genomic)  
(iii) HYPOTHETICAL: NO  
(iv) ANTI-SENSE: YES  
(xi) SEQUENCE DESCRIPTION: SEQ ID NO:85:  
20 CTCCATGTCC CAACTCCGAT CAC 23  
(2) INFORMATION FOR SEQ ID NO:86:  
(i) SEQUENCE CHARACTERISTICS:  
(A) LENGTH: 38 base pairs  
(B) TYPE: nucleic acid  
25 (C) STRANDEDNESS: single  
(D) TOPOLOGY: unknown  
(ii) MOLECULE TYPE: DNA (genomic)  
(iii) HYPOTHETICAL: NO  
(iv) ANTI-SENSE: NO  
30 (xi) SEQUENCE DESCRIPTION: SEQ ID NO:86:  
GGTTTTCCCA GTCACGACGT TGTAACACGA CGGCCAGT 38  
(2) INFORMATION FOR SEQ ID NO:87:  
(i) SEQUENCE CHARACTERISTICS:  
(A) LENGTH: 38 base pairs  
35 (B) TYPE: nucleic acid  
(C) STRANDEDNESS: single  
(D) TOPOLOGY: unknown  
(ii) MOLECULE TYPE: DNA (genomic)  
40 (iii) HYPOTHETICAL: NO  
(iv) ANTI-SENSE: NO  
(xi) SEQUENCE DESCRIPTION: SEQ ID NO:87:  
GGUUUCCCA GUCACGACGU UGUAAAACGA CGGCCAGU 38  
(2) INFORMATION FOR SEQ ID NO:88:  
(i) SEQUENCE CHARACTERISTICS:  
45 (A) LENGTH: 35 base pairs  
(B) TYPE: nucleic acid  
(C) STRANDEDNESS: single  
(D) TOPOLOGY: unknown  
(ii) MOLECULE TYPE: DNA (genomic)  
50 (iii) HYPOTHETICAL: NO  
(iv) ANTI-SENSE: NO  
(xi) SEQUENCE DESCRIPTION: SEQ ID NO:88:  
GACGACGACA AGATGCCCTG CTCTGAAGAG ACACC 35  
(2) INFORMATION FOR SEQ ID NO:89:

-103-

## (i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 35 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: unknown

(ii) MOLECULE TYPE: DNA (genomic)

(iii) HYPOTHETICAL: NO

(iv) ANTI-SENSE: YES

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:89:

GGAACAAGAC CCGTTTAATT CTTTCCAGTG GAACC

35